Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles

Thomas G. Graeber[1] & David Eisenberg[1]

Published online: 29 October 2001, DOI: 10.1038/ng755

Many biological signaling pathways involve autocrine ligand–receptor loops; misregulation of these signaling loops can contribute to cancer phenotypes. Here we present an algorithm for detecting such loops from gene expression profiles. Our method is based on the hypothesis that for some autocrine pathways, the ligand and receptor are regulated by coupled mechanisms at the level of transcription, and thus ligand–receptor pairs comprising such a loop should have correlated mRNA expression. Using our database of experimentally known ligand–receptor signaling partners, we found examples of ligand–receptor pairs with significantly correlated expression in five cancer-based gene expression datasets. The correlated ligand–receptor pairs we identified are consistent with known autocrine signaling events in cancer cells. In addition, our algorithm predicts new autocrine signaling loops that can be verified experimentally. Chemokines were commonly members of these potential autocrine pathways. Our analysis also revealed ligand–receptor pairs with expression patterns that may indicate cellular mechanisms for preventing autocrine signaling.

Autocrine signaling (production of, and response to, a ligand by the same cell) is a common mechanism of signal transduction in normal physiological processes[1]. The effects of autocrine signaling can also be amplified by paracrine signaling between neighboring homotypic cells[1]. During tumorigenesis, misregulated autocrine signaling can result in cancer cells less dependent on survival and growth factors from surrounding tissues[1,2].

Correlated expression of two genes at the mRNA level suggests that the genes are regulated by a common underlying mechanism and that their products may function together[3–6]. As autocrine signaling relies upon the expression of the appropriate ligand and receptor by the same cell, we reasoned that for some autocrine pathways, expression of the ligand and receptor involved would be coupled at the transcriptional level. We attempted to identify known and candidate autocrine ligand–receptor loops by systematically searching gene expression data for ligand–receptor partners that show correlated expression.

Our algorithm integrates two types of data: experimentally determined ligand–receptor cognate pairs and measured gene expression profiles. We first compiled the Database of Ligand–Receptor Partners (DLRP) from published papers and reviews. This database includes cytokines, chemokines, and growth, angiogenesis and developmental factors, and contains 175 protein ligands, 131 protein receptors and 451 experimentally determined ligand-receptor pairings. The contents of DLRP are available through the Database of Interacting Proteins (DIP; http://dip.doe-mbi.ucla.edu/)[7]. For gene expression data, we gathered five publicly available datasets from primary-source web sites (Methods)[8–12]. Each of these datasets contains gene expression values for thousands of genes in approximately 50 patient cancer samples (Table 1).

A schematic of our approach is shown in Fig. 1. For each ligand–receptor pair measured in the expression datasets, we calculated the Pearson correlation coefficient (r) between the ligand and receptor expression profiles. We then calculated two probabilities for each correlation coefficient: the probability of obtaining a larger (in absolute value) correlation coefficient by randomly shuffling one of the expression profiles (P_{shuffle}), and the probability of finding any other gene pair within the dataset with a higher absolute correlation coefficient ($P_{\text{all pairs}}$). The ligand–receptor partners with the most significantly correlated expression ($P_{\text{shuffle}} \leq 0.002$ and $P_{\text{all pairs}} \leq 0.06$) in the diffuse large B-cell lymphoma (DLBCL) and leukemia datasets are shown in Tables 2 and 3 (for all five datasets, see Web Tables A–E). The most correlated ligand–receptor pairs show correlation similar to

Table 1 • Gene expression datasets used in our analysis

Expression profile dataset	Sample descriptions and subsets (n*)		Number of ligands	Number of receptors	Number of receptor–ligand pairs	Array type
DLBCL[8]	DLBCL (46) surviving (18)	non-surviving (22)	45	45	37	cDNA
leukemia[9]	AML (25)	ALL (47)	66	62	68	oligo
breast[10]	breast cancer (62)	breast cell lines and normal samples (22)	33	38	33	cDNA
NCI60[11]	cell lines from a panel of various cancer types (64)		37	49	45	cDNA
colon[12]	colon cancer (40)	colon normal (22)	54	64	66	oligo

*number of patient and cell line samples in each dataset or subset.

[1]*Howard Hughes Medical Institute, UCLA–Department of Energy Laboratory of Structural Biology and Molecular Medicine, Departments of Chemistry and Biochemistry and Biological Chemistry, University of California, Los Angeles, California 90095, USA. Correspondence should be addressed to D.E. (e-mail: david@mbi.ucla.edu).*

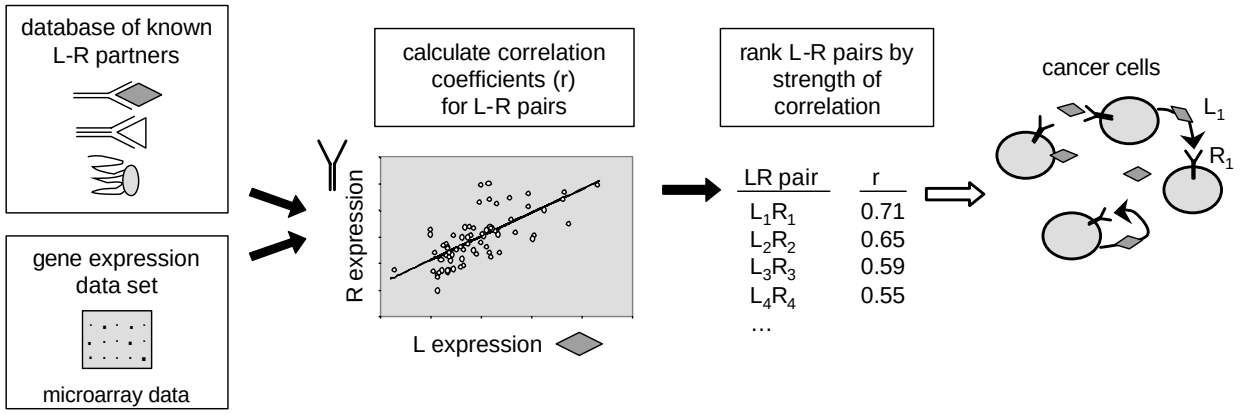

Fig. 1 Bioinformatic identification of potential autocrine ligand–receptor signaling loops. For each known ligand (L)–receptor (R) pair measured in a particular gene expression dataset, we determined the correlation coefficient (*r*) between the L and R expression profiles. L–R pairs with high correlation coefficients are candidate autocrine signaling mediators in the cancer subtypes that show L–R coexpression.

that of gene pairs encoding proteins that are known to function together, such as ribosomal proteins or protein pairs from the Database of Interacting Proteins (Web Figs. A and B). Considered as one group, ligand–receptor cognate pairs tend to have more correlated gene pairs than do random groups of gene pairs, but this tendency is not statistically significant.

To assess the usefulness of our approach, we first determined whether it identified any ligand–receptor partners previously implicated in autocrine signaling. Of the 37 most significantly correlated ligand–receptor pairs ($P_{shuffle} \leq 0.002$ and $P_{all\ pairs} \leq 0.06$) from all five datasets, eight pairs have previously been implicated in autocrine signaling in tumor cells (Tables 2 and 3 and Web Tables A–E). For example, the interleukin-10 (IL10) ligand and its receptor (IL10RA) show correlated expression in the acute myelogenous leukemia (AML) samples ($P_{all\ pairs} < 0.02$; Fig. 2 and Table 3). Previous studies have reported direct evidence for the involvement of IL10 and IL10RA in autocrine signaling in many cell types, including monocytes and leukemia cells[13,14]. Notably, IL10 and IL10RA are not correlated in the acute lymphoblastic leukemia (ALL) samples (Fig. 2).

Another example of a previously reported autocrine pathway identified by our algorithm involves the tumor necrosis factor (TNF) ligand and its receptors, TNFRSF1A (TNFR1) and TNFRSF1B (TNFR2). In the lymphoma (DLBCL) dataset, we separately analyzed samples from patients who did, or did not, survive their disease beyond four years after diagnosis and standard chemotherapy regimens[8]. The subgroup of patients who did not survive beyond four years showed correlated expression between TNF and TNFR1, and both subgroups had correlated expression between TNF and TNFR2 (Table 2). A TNF autocrine loop contributing to cellular proliferation has previously been observed in lymphoma cell lines[15,16].

The overlap between our results and previous reports of autocrine signaling supports our premise that for some ligand–receptor pairs involved in autocrine signaling, the ligand–receptor co-function is reflected at the mRNA expression level. We predict that the remaining correlated ligand–receptor pairs from our results are also involved in autocrine signaling.

Examination of the ligand–receptor pairs with correlated expression shows several trends. For one, ligand–receptor pairs from the CC family of chemokines are significantly over-represented. We found CC chemokine ligand–receptor pairs with correlated expression in the DLBCL, leukemia, breast cancer and colon datasets (Tables 2 and 3; Web Tables A–C, E). For example, the CCL4/MIP-1β ligand and its receptor CCR5 show correlated expression in both the DLBCL ($P_{all\ pairs} < 0.002$) and leukemia ($P_{all\ pairs} < 0.01$) datasets (Tables 2 and 3). In the DLBCL and leukemia datasets, CC chemokine ligand–receptor pairs make up less than 20% of the ligand–receptor pairs measured but constitute 50% or more of the co-expressed ligand–receptor pairs with

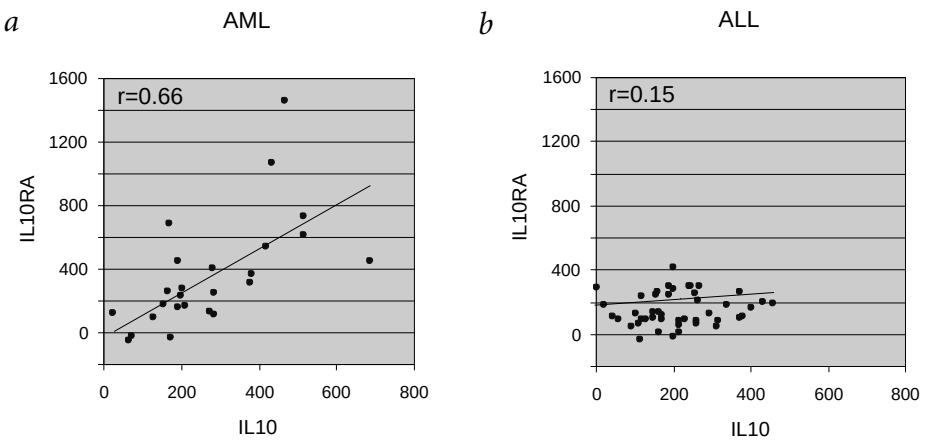

Fig. 2 Identification of a previously reported autocrine signaling loop. Using the method shown in Fig. 1, we assessed the potential of the IL10 ligand and its receptor, IL10RA, to form an autocrine loop. Data points represent individual samples from individuals with leukemia. Absolute amounts of mRNA measured using oligo-based gene expression arrays are reported on each axis. Correlation coefficients (*r*) are indicated, and the lines show least-squares linear regression fits to the data. ***a***, Correlated expression of IL10 and IL10RA in individuals with AML ($P_{all\ pairs} < 0.02$; Table 3). ***b***, Lack of correlation between IL10 and IL10RA expression in individuals with ALL. Previous reports have shown autocrine signaling through IL10 and IL10RA in monocytes.

Table 2 • Correlated ligand–receptor pairs from the lymphoma (DLBCL) dataset

Ligand	Receptor	Correlation coefficient (*r*)			$P_{shuffle}$			$P_{all\ pairs}$		
		Surviving	Nonsurviving	All	Surviving	Nonsurviving	All	Surviving	Nonsurviving	All
		n=18	*n*=22	*n*=46						
CCL5[a] (RANTES)	CCR5	0.78 0.45	0.75 0.57	**0.79** 0.60	0.0018 0.070	0.0020 0.018	6.8×10^{-7} 0.00012	0.0012 0.11	0.0035 0.045	0.00032 0.0065
CCL5[a] (RANTES)	CCR1	0.45 0.58	0.66 0.49	**0.74** 0.61	0.063 0.017	0.0041 0.033	1.6×10^{-6} 7.5×10^{-5}	0.11 0.033	0.015 0.098	0.00069 0.0058
CCL4[b] (MIP-1β)	CCR1	0.57	0.63	**0.68**	0.019	0.0058	1.2×10^{-5}	0.037	0.022	0.0019
CCL4[c,d] (MIP-1β)	CCR5	0.53	0.64	**0.68**	0.033	0.0088	1.8×10^{-5}	0.055	0.021	0.0019
CCL11[b] (eotaxin)	CCR5	0.51	0.57	**0.59**	0.043	0.021	0.00022	0.070	0.044	0.0085
TNF[a,d]	TNFRSF1A (TNFR1)	–0.13 0.14 –0.07	**0.71** 0.69 0.64	0.42 0.46 0.31	0.60 0.55 0.77	0.0015 0.0021 0.0055	0.0055 0.0021 0.040	0.66 0.64 0.82	0.0070 0.010 0.021	0.071 0.048 0.20
TNF[a,d]	TNFRSF1B (TNFR2)	0.46 0.56 0.39	0.50 0.58 0.46	0.35 **0.42** 0.23	0.070 0.018 0.10	0.026 0.010 0.045	0.026 0.0064 0.13	0.11 0.043 0.18	0.091 0.043 0.12	0.15 0.075 0.33

Individuals with DLBCL were divided into two subgroups based on whether they did, or did not, survive their disease beyond 4 years after diagnosis and chemotherapy[8]. All results with $P_{shuffle}\leq0.002$ and $P_{all\ pairs}\leq0.06$ are shown (Methods). Results involving TNFRSF1B were also included for comparison to the TNFRSF1A results. Complete results and sequence identifiers of all genes, for this dataset and others, are available in Supplementary Information. Bold values indicate the most significant correlation coefficient for each ligand–receptor pair based on $P_{shuffle}$. *n*, number of patient samples in each dataset or subset. [a]Genes that were measured either in replicate or using multiple sequence isoforms. [b]Ligands that are poor agonists of or have weak affinity for the indicated receptor[26]. CCL11/Eotaxin also binds CCR3, which was not measured in the DLBCL dataset. [c]Ligand–receptor pairs displaying correlated expression in multiple datasets. [d]Ligands previously observed in autocrine loops[15–17].

$P_{shuffle}\leq0.002$ and $P_{all\ pairs}\leq0.06$. For both of these datasets, the probability of the observed enrichment occurring by chance is less than 5%. The CC chemokines CCL2/MCP-1, CCL3/MIP-1α and CCL4 have previously been shown to function in autocrine pathways[17,18]. Our results support previous evidence that chemokine signaling mechanisms contribute to tumor proliferation, mobility and invasiveness[19,20] and can be used to guide research in this field.

The abundance of CC chemokine ligand–receptor pairs with correlated expression prompted further investigation into potential biological roles for CC chemokines in these cancer samples. Chemokines function in chemotaxis to stimulate and guide the movement of cells; accordingly, chemokine signaling modulates cellular adhesion[21]. Exposure of responsive cells to CC chemokines causes both increased surface expression of β2 integrins and an increased affinity of β2 integrins for adhesion molecules[21–23]. We found that the β2 integrins ITGAX/CD11c and ITGB2/CD18 and the β5 integrin ITGB5, but not ITGAL/CD11a or other integrins, were highly expressed when the CC chemokine ligand–receptor pairs from Table 2 were present in DLBCLs (Fig. 3). These results

Table 3 • Correlated ligand–receptor pairs from the leukemia (AML/ALL) dataset

Ligand	Receptor	Correlation coefficient (*r*)			$P_{shuffle}$			$P_{all\ pairs}$		
		AML	ALL	AML/ALL	AML	ALL	AML/ALL	AML	ALL	AML/ALL
		n=25	*n*=47	*n*=72						
CCL2[a,b] (MCP-1)	FY[c] (DARC)	0.67 0.51	0.68 0.50	**0.70** 0.54	0.0011 0.014	4.4×10^{-6} 0.00065	3.2×10^{-9} 5.7×10^{-6}	0.012 0.073	0.0058 0.055	0.0014 0.022
CCL4[a,b,d] (MIP-1β)	CCR5	0.65 0.38	0.63 0.63	**0.61** 0.57	0.0029 0.078	2.6×10^{-5} 2.1×10^{-5}	6.6×10^{-7} 3.3×10^{-6}	0.016 0.19	0.013 0.011	0.0086 0.016
CCL4′		0.44	0.20	0.29	0.031	0.18	0.013	0.13	0.45	0.21
FGF4	FGFR2[a]	0.53 0.30 0.36 –0.22	0.57 0.45 0.34 –0.09	**0.54** 0.40 0.26 –0.17	0.010 0.15 0.084 0.29	0.00013 0.0024 0.025 0.54	7.0×10^{-6} 0.00079 0.031 0.16	0.059 0.31 0.22 0.46	0.027 0.087 0.21 0.73	0.022 0.092 0.27 0.47
EFNB1	EPHB4	0.66	0.46	**0.51**	0.0012	0.0019	1.7×10^{-5}	0.013	0.084	0.032
IL10[d,e]	IL10RA	0.66	0.15	**0.48**	0.0019	0.30	7.1×10^{-5}	0.013	0.55	0.043
CCL8 (MCP-2)	CCR5	0.51	0.47	**0.47**	0.014	0.0015	7.9×10^{-5}	0.072	0.076	0.048
CCL7[c,e] (MCP-3)	CCR5	0.01	**0.49**	0.30	0.98	0.00099	0.012	0.99	0.060	0.21
JAG1[e]	NOTCH4	**0.65**	0.13	0.34	0.0014	0.38	0.0040	0.016	0.62	0.15

All results with $P_{shuffle}\leq0.002$ and $P_{all\ pairs}\leq0.06$ are shown (Methods; see also Table 2, legend). [a]Genes that were measured either in replicate or using multiple sequence isoforms. CCL4′ is an isoform of CCL4 with an 8-nt gap and a 1-nt change. The FGFR2 sequences consist of different splice variants. [b]Ligand–receptor pairs showing correlated expression in multiple datasets. [c]FY/DARC is a nonsignaling chemokine receptor, and CCL7/MCP-3 is an antagonist for CCR5[26]. [d]Ligands previously observed in autocrine loops[13,14,17]. [e]AML- or ALL-specific results. *n*, number of patient samples in each dataset or subset.

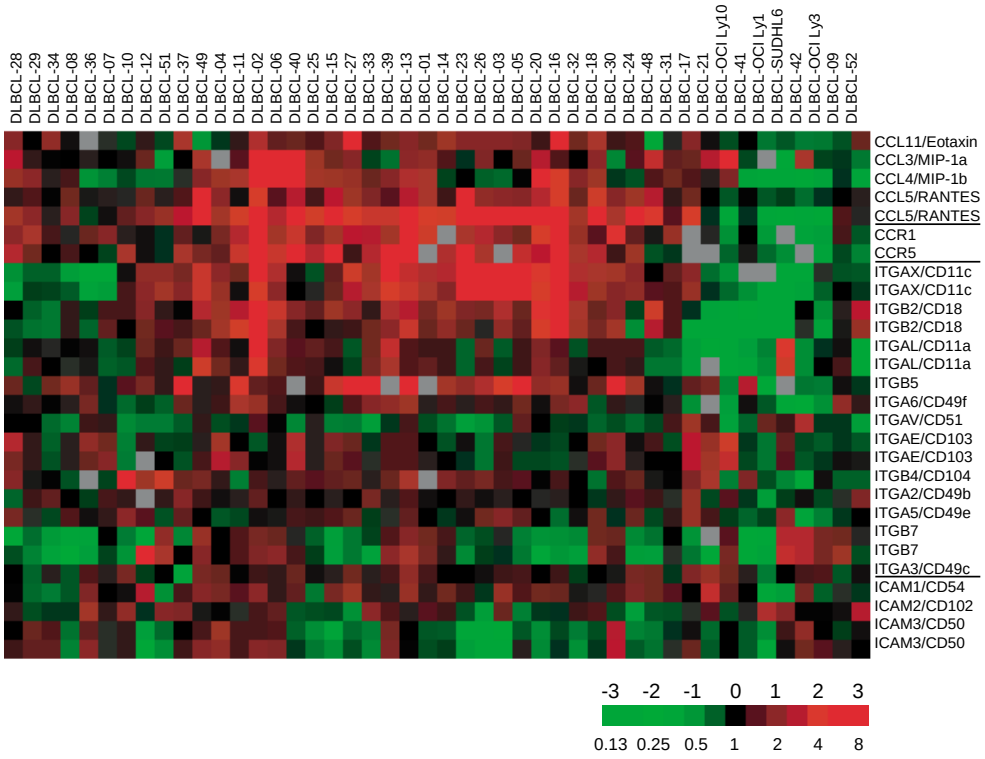

Fig. 3 Correlated expression of CC chemokines, their receptors and integrins in DLBCLs. Note the similar expression profiles for the CC chemokine ligand–receptor pairs and a subset of the β2 integrins: ITGAX/CD11c and ITGB2/CD18, but not ITGAL/CD11a. These results agree with the spectrum of integrins previously seen to have induced cell-surface expression after exposure to CC chemokines. Each row shows the expression profile of a gene across multiple DLBCL samples, with each column representing a different sample. Expression levels are reported as the ratio of the expression level of the experimental mRNA sample relative to a reference mRNA sample, measured using cDNA-based gene expression arrays. Expression is indicated according to the color scale shown, with green indicating low relative expression and red indicating high expression. The scale extends from ratios of 0.13–8 (–3 to 3 in log base 2 units). We created this figure using the programs Cluster and Treeview[5]. The genes are grouped in gene families separated by solid lines: CC chemokine ligands (CCLs), CC chemokine receptors (CCRs), β2 integrins (ITGAL/CD11a, ITGAX and ITGB2), integrin β5 (ITGB5) and intercellular adhesion molecules (ICAMs). We observed similar results in the breast cancer dataset (data not shown).

agree with the spectrum of integrins previously observed to have induced cell-surface expression after exposure to the CC chemokines CCL2, CCL3 and CCL5/RANTES[17,22,23]. Thus, these three classes of proteins, CC chemokines, their receptors and β2 integrins, are probably regulated by coupled mechanisms in these cell types. As integrins have been implicated in both lymphoma aggressiveness[24] and invasiveness[19,25], the subset of DLBCLs that co-express chemokine ligand–receptor pairs and integrins may show distinct phenotypic characteristics.

In addition to identifying potential autocrine pathways, our analysis revealed ligand–receptor pairs with expression patterns that may indicate mechanisms for preventing autocrine signaling. For example, the CCL2 ligand and FY/DARC receptor are co-expressed in colon samples ($P_{\text{all pairs}}$<0.003), leukemias ($P_{\text{all pairs}}$<0.02) and breast cancers ($P_{\text{all pairs}}$<0.12; Table 3 and Web Tables B,C and E). To date, the FY receptor has not been found to generate intracellular signaling upon ligand binding[26], although CCL2 and other FY-binding chemokines can bind to other chemokine receptors that do signal. FY thus seems to modulate chemokine signaling by acting as a chemokine sink (decoy receptor)[27]. Our observation that the nonsignaling FY receptor is co-expressed in cells expressing the CCL2 ligand may indicate a mechanism for reducing the response of a cell to its own CCL2 secretion.

We also observed examples of negatively correlated expression of ligand–receptor pairs, which may reflect an underlying mechanism to prevent autocrine signaling by reducing the level of a receptor whenever its cognate ligand is produced. The growth factor neuregulin-1 (NRG1) and its receptor ERBB3/HER3 show negatively correlated expression in multiple cancer types in the NCI60 ($P_{\text{all pairs}}$<0.03), breast cancer ($P_{\text{all pairs}}$<0.05) and DLBCL ($P_{\text{all pairs}}$<0.1) datasets (Web Tables A, C and D). The predominant receptor complex for the NRG1 ligand is a heterodimeric receptor consisting of ERBB3 and ERBB2 (ERBB2 is commonly overexpressed in breast cancer through gene amplification)[28]. Whereas ERBB3 expression is negatively correlated with NRG1 expression in these datasets, ERBB2 expression is not correlated either positively or negatively with NRG1 expression (Web Tables A, C and D). As ERBB2 and ERBB3 receptors cannot efficiently generate NRG1-stimulated signaling without forming heterodimeric complexes with each other[28], modulation of ERBB3 levels could be sufficient to regulate the sensitivity of a cell to the NRG1 ligand. Thus, negatively correlated expression of NRG1 and ERBB3 may reflect a cellular mechanism for preventing an autocrine response to NRG1. Other examples of negatively correlated ligand–receptor expression can be found in Web Tables A–E.

One major obstacle to the correct interpretation of gene expression data from human tumor samples is the potential for the samples to contain contaminating cells from other tissue types[10]. If our findings are due to contamination by other cell types, this would mean that rather than detecting potential autocrine loops within the tumors, we are instead detecting either potential autocrine loops within the contaminating cell population, or paracrine loops between the tumor cells and the contaminating cells. Nevertheless, such signaling loops could still be relevant to the biology of the tumor. For the potential autocrine or paracrine loops we have identified, gene expression data from tissue arrays or histology slides could be used to resolve the cell types expressing the ligand and receptor.

A second limitation of gene expression analysis comes from the observation that protein and mRNA expression levels for a gene are not always well correlated[29]. The results presented here are

based upon measurements of mRNA expression levels. But receptors and ligands must be expressed as proteins and either displayed extracellularly or secreted in order to be functional. In general, the role of post-transcriptional modulation of receptors and ligands must also be considered.

Because our method relies on detectable differences in expression, it may be biased toward identifying ligand–receptor pairs from gene families that are strongly modulated between cell activation states or cell types. This may partially explain the abundance of chemokine and cytokine ligand–receptor pairs identified by our analysis.

We have based our analysis on the observation that coordinated expression suggests co-function[4,6]. Nevertheless, the candidate ligand–receptor pairs identified here must be experimentally tested to determine if they truly participate in biologically relevant autocrine signaling loops. Compared with randomly chosen ligand–receptor cognate pairs, the candidate ligand–receptor pairs identified here should more often survive experimental validation of autocrine function.

Rather than predicting which gene products function together, our method identifies the particular cancer types in which known ligand–receptor pairs probably function together. A variation of our approach could also be used to identify candidate genes as new partners for established, or orphan, ligands and receptors; for example, by using expression data of previously uncharacterized genes or ESTs. Our approach could also be used to search for candidate paracrine signaling pathways between two different tissue types.

Methods

Gene expression datasets. We obtained the following datasets from the sources indicated: DLBCL dataset, Lymphoma and Leukemia Molecular Profiling Project (LLMPP), National Cancer Institute (NCI; http://llmpp.nih.gov/)[8]; leukemia dataset, Molecular Pattern Recognition, Whitehead/MIT Center for Genome Research (http://waldo.wi.mit.edu/MPR/)[9]; breast dataset, Molecular Portraits, Stanford Microarray Database (http://genome-www.stanford.edu/)[10]; NCI60 dataset, NCI60 Cancer Microarray Project, Stanford Microarray Database (http://genome-www.stanford.edu/)[11]; colon dataset, Princeton University Gene Expression Project (http://microarray.princeton.edu/oncology/)[12].

These datasets were generated by gene microarray–based measurements of mRNA expression of thousands of genes in patient cancer samples, normal tissue samples and cell lines (Table 1). The DLBCL, breast and NCI60 datasets were acquired using cDNA-based gene expression arrays; the expression values measured are ratios of an experimental mRNA sample relative to a reference mRNA sample mixture[8]. The leukemia and colon datasets were generated using oligo-based gene expression arrays; the expression values are absolute measurements of mRNA levels. Because the datasets were collected using different types of gene expression arrays, with different sets of genes measured, we analyzed them individually. Genes with expression levels that were too low to be reliably measured were not included in the analysis[8–12]. Each gene expression dataset contains expression measurements for only a subset of the ligand–receptor pairs in our Database of Ligand–Receptor Partners (Table 1).

Correlation measure. We measured correlations between gene expression profiles using the standard Pearson correlation coefficient

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}},$$

where x_i and y_i represent the amounts of mRNA for the ligand and receptor in sample i and $\bar{x}$ and $\bar{y}$ indicate the respective mean expression values[5,30]. Before calculating correlation coefficients, we eliminated any expression measurement of a gene that was more than four s.d. from the mean expression level of the gene. We determined this cutoff empirically and used it to eliminate outliers. We also plotted expression comparisons and visually inspected them for outliers. We calculated correlation measurements of cDNA-based profiles using log base 2–transformed data, as in the original papers[8,10,11]. We also verified that untransformed data yield similar results.

If the patient samples in an expression dataset could be divided into subcategories (such as AML and ALL), then for each ligand–receptor pair we calculated correlation coefficients for each subset and for the set as a whole. When we observed more than one version of the same gene on a DNA microarray, we considered each measurement separately and then grouped them together (Tables 2 and 3 and Web Tables A–E). In general, multiple measurements of the same gene agreed; when they did not, the difference could usually be attributed to the use of different isoforms (for instance, splice variants) of the same genes (for known sequence variants, see Table 3, legend). For both cDNA-based and oligo-based gene expression arrays, however, the sequence information for the genes used is incomplete. As a result, some inconsistencies could not be definitively resolved. In addition, some inconsistencies could be related to the many technical issues involved in array manufacturing, hybridization and scanning. We include inconsistent results here (Tables 2 and 3 and Web Tables A–E) for their potential utility in guiding subsequent studies.

Determination of probabilities. We determined the probability of obtaining a larger (in absolute value) correlation at random between genes A and B (P_{shuffle}) by randomly shuffling the expression profile of gene A and then determining the correlation coefficient between the shuffled profile and the original profile of gene B. We did this 100,000 times; the resulting distribution was fit to a Gaussian density function. We then determined the probability of obtaining the observed correlation coefficient between gene A and B using the complement error function[30]. This shuffling method preserves the properties (such as mean, variance and dimensionality) of each expression profile during the randomization. We also determined the probability of finding any other pair of genes within the dataset with a higher absolute correlation coefficient ($P_{\text{all pairs}}$), and the probability of finding another ligand–receptor pair, including those that do not bind each other, with a higher absolute correlation coefficient ($P_{\text{all LR pairs}}$) after calculating the distributions of all correlation coefficients between all pairs of genes for each set and subset. There was no significant difference between $P_{\text{all pairs}}$ and $P_{\text{all LR pairs}}$.

Note: Supplementary information is available on the Nature Genetics web site (http://genetics.nature.com/supplementary_info/).

Acknowledgments

This paper is dedicated to the memory of Nathan B. Friedman, his inspiring spirit, and his contributions to pathology. We thank the authors of the gene expression papers cited here for making their data publicly available. We also thank M. Balbirnie, R. Grothe, G. Kleiger, R. Landgraf, P. Mallick, E. Marcotte, M. Pellegrini, L. Salwinski, and I. Xenarios for insights and helpful discussions. This work was supported by grants from the US Department of Energy (DOE) and National Institutes of Health. T.G.G. was supported by an Alfred P. Sloan Foundation/DOE postdoctoral fellowship.

Received 24 May; accepted 28 September 2001.

1. Dawson, T. & Wynford-Thomas, D. Does autocrine growth factor secretion form part of a mechanism which paradoxically protects against tumour development? *Br. J. Cancer* **71**, 1136–1141 (1995).
2. Sporn, M.B. & Roberts, A.B. Autocrine growth factors and cancer. *Nature* **313**, 745–747 (1985).
3. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. & Church, G.M. Systematic determination of genetic network architecture. *Nature Genet.* **22**, 281–285 (1999).
4. DeRisi, J.L., Iyer, V.R. & Brown, P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
5. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
6. Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
7. Xenarios, I. *et al.* DIP: The Database of Interacting Proteins: 2001 update. *Nucleic Acids Res.* **29**, 239–241 (2001).
8. Alizadeh, A.A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
9. Golub, T.R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
10. Perou, C.M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**,

747–752 (2000).
11. Ross, D.T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet.* **24**, 227–235 (2000).
12. Alon, U. *et al.* Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **96**, 6745–6750 (1999).
13. Capsoni, F. *et al.* Development of phagocytic function of cultured human monocytes is regulated by cell surface IL-10. *Cell. Immunol.* **189**, 51–59 (1998).
14. Peng, B., Mehta, N.H., Fernandes, H., Chou, C.C. & Raveche, E. Growth inhibition of malignant CD5+B (B-1) cells by antisense IL-10 oligonucleotide. *Leuk. Res.* **19**, 159–167 (1995).
15. Giri, D.K. & Aggarwal, B.B. Constitutive activation of NF-κB causes resistance to apoptosis in human cutaneous T cell lymphoma HuT-78 cells. Autocrine role of tumor necrosis factor and reactive oxygen intermediates. *J. Biol. Chem.* **273**, 14008–14014 (1998).
16. Warzocha, K. *et al.* Plasma levels of tumour necrosis factor and its soluble receptors correlate with clinical features and outcome of Hodgkin's disease patients. *Br. J. Cancer* **77**, 2357–2362 (1998).
17. Tanaka, Y. *et al.* Constitutive chemokine production results in activation of leukocyte function-associated antigen-1 on adult T-cell leukemia cells. *Blood* **91**, 3909–3919 (1998).
18. Zhou, P., Thomassen, M.J., Pettay, J., Deodhar, S.D. & Barna, B.P. Human monocytes produce monocyte chemoattractant protein 1 (MCP-1) in response to a synthetic peptide derived from C-reactive protein. *Clin. Immunol. Immunopathol.* **74**, 84–88 (1995).
19. Müller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
20. Wang, J.M., Deng, X., Gong, W. & Su, S. Chemokines and their role in tumor growth and metastasis. *J. Immunol. Methods* **220**, 1–17 (1998).
21. Carr, M.W., Alon, R. & Springer, T.A. The C-C chemokine MCP-1 differentially modulates the avidity of β1 and β2 integrins on T lymphocytes. *Immunity* **4**, 179–187 (1996).
22. Vaddi, K. & Newton, R.C. Regulation of monocyte integrin expression by β-family chemokines. *J. Immunol.* **153**, 4721–4732 (1994).
23. Jiang, Y., Beller, D.I., Frendl, G. & Graves, D.T. Monocyte chemoattractant protein-1 regulates adhesion molecule expression and cytokine production in human monocytes. *J. Immunol.* **148**, 2423–2428 (1992).
24. Erikstein, B.K. *et al.* Expression of CD18 (integrin β2 chain) correlates with prognosis in malignant B cell lymphomas. *Br. J. Haematol.* **83**, 392–398 (1993).
25. Stroeken, P.J., van Rijthoven, E.A., van der Valk, M.A. & Roos, E. Targeted disruption of the β1 integrin gene in a lymphoma cell line greatly reduces metastatic capacity. *Cancer Res.* **58**, 1569–1577 (1998).
26. Murphy, P.M. *et al.* International union of pharmacology. XXII. Nomenclature for chemokine receptors. *Pharmacol. Rev.* **52**, 145–176 (2000).
27. Dawson, T.C. *et al.* Exaggerated response to endotoxin in mice lacking the Duffy antigen/receptor for chemokines (DARC). *Blood* **96**, 1681–1684 (2000).
28. Sliwkowski, M.X. *et al.* Co-expression of erbB2 and erbB3 proteins reconstitutes a high affinity receptor for heregulin. *J. Biol. Chem.* **269**, 14661–14665 (1994).
29. Pradet-Balade, B., Boulmé, F., Beug, H., Müllner, E.W. & Garcia-Sanz, J.A. Translation control: bridging the gap between genomics and proteomics? *Trends. Biochem. Sci.* **26**, 225–229 (2001).
30. Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. *Numerical Recipes in C* (Cambridge University Press, Cambridge, UK, 1992).