

Research article

Open Access

A reference database for tumor-related genes co-expressed with interleukin-8 using genome-scale *in silico* analysis

Lawrence Benbow¹, Lynn Wang², Maureen Laverty¹, Suxing Liu², Ping Qiu¹, Richard W Bond², Eric Gustafson¹, Joseph A Hedrick¹, Mitchell Kostich¹, Jonathan R Greene¹ and Luquan Wang*¹

Address: ¹Discovery Technology Department, Schering-Plough Research Institute, 2015 Galloping Hill Rd, Kenilworth, NJ 07033, USA and ²Tumor Biology Department, Schering-Plough Research Institute, 2015 Galloping Hill Rd, Kenilworth, NJ 07033, USA

E-mail: Lawrence Benbow - lawrence.benbow@spcorp.com; Lynn Wang - lynn.wang@spcorp.com; Maureen Laverty - maureen.laverty@spcorp.com; Suxing Liu - suxing.liu@spcorp.com; Ping Qiu - ping.qiu@spcorp.com; Richard W Bond - richard.bond@spcorp.com; Eric Gustafson - eric.gustafson@spcorp.com; Joseph A Hedrick - joseph.hedrick@spcorp.com; Mitchell Kostich - mitchell.kostich@spcorp.com; Jonathan R Greene - jonathan.greene@spcorp.com; Luquan Wang* - luquan.wang@spcorp.com

*Corresponding author

Published: 11 October 2002

Received: 28 August 2002

BMC Genomics 2002, 3:29

Accepted: 11 October 2002

This article is available from: <http://www.biomedcentral.com/1471-2164/3/29>

© 2002 Benbow et al; licensee BioMed Central Ltd. This article is published in Open Access: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The EST database provides a rich resource for gene discovery and *in silico* expression analysis. We report a novel computational approach to identify co-expressed genes using EST database, and its application to IL-8.

Results: IL-8 is represented in 53 dbEST cDNA libraries. We calculated the frequency of occurrence of all the genes represented in these cDNA libraries, and ranked the candidates based on a Z-score. Additional analysis suggests that most IL-8 related genes are differentially expressed between non-tumor and tumor tissues. To focus on IL-8's function in tumor tissues, we further analyzed and ranked the genes in 16 IL-8 related tumor libraries.

Conclusions: This method generated a reference database for genes co-expressed with IL-8 and could facilitate further characterization of functional association among genes.

Background

Chemokines are a large superfamily of small, structurally-related peptides originally discovered as neutrophil attractants. Interleukin-8 (IL-8) is a potent member of the supergene family of CXC chemokines with ELR motif (ELR+). ELR+ CXC chemokines are potent angiogenic factors, whereas ELR- CXC chemokines are potent angiostatic factors. Studies have shown that these angiogenesis-related activities are correlated with tumorigenesis in many tu-

mor types and are distinct from their ability to recruit neutrophils [1,2].

Interleukin-8 (IL-8) is inducible in a wide range of cells including lymphocytes, monocytes, endothelial cells, fibroblasts, hepatocytes, and keratinocytes [3-5]. IL-8 is also found to be constitutively expressed in several tumor tissues including bronchogenic carcinoma, non-small cell lung cancer, colorectal carcinoma, breast cancer, melanoma, prostate cancer, gastric carcinoma, and ovarian can-

Table 1: dbEST cDNA library summary. The cDNA libraries are catalogued into libraries where IL-8, GRO- α , or GRO- β is represented. Tumor libraries are cDNA libraries prepared from tumor tissues or tumor cell lines based on dbEST library annotation. See Additional file 1 for the detailed description of cDNA libraries where IL-8 is represented.

Gene(s)	Total Libraries			Tumor Libraries		
	ESTs	Clones	Libraries	ESTs	Clones	Libraries
IL-8	329649	306888	53	173152	171521	16
GRO- α	337364	349476	34	175968	174072	13
GRO- β	224993	213645	27	161390	159808	12
dbEST Total	3350920	3043217	4419	896990	864427	2734

cer. While IL-8 has been implicated in growth-potential [6], angiogenesis [7], metastasis [3,8], and tumorigenesis [3,6] of various tumors, its specific role remains unclear. Several observations support the assertion that a complex interaction between IL-8 and several other growth factors, cytokines, or other proteins is responsible for these tumor-related events [5,9–13]. Growth regulated protein alpha (GRO- α), beta (GRO- β), and ENA-78 have been reported to be co-induced with IL-8 in A549 cells stimulated with two proinflammatory cytokines, IL-1 β and TNF- α [14]. This result is consistent with the presence of nuclear factor kappa B (NF- κ B) consensus binding sites in the promoter regions of all three genes [4,15]. In addition, GRO- α and - β are also members of the ELR⁺ CXC cytokine family that are reported to be important mediators of tumorigenesis through their angiogenic properties [2,16]. They both share one of the IL-8 receptors, CXCR2, which has been postulated to regulate the ELR⁺ CXC chemokine-mediated angiogenesis and resulting-tumorigenesis [1]. These observations suggest that GRO- α and - β may exhibit expression profiles similar to IL-8. Coordinated expression of these and other factors with IL-8 in certain tumor tissues suggests that they may be functionally associated with IL-8 in these tumors. In order to fully understand IL-8's role in these events, it is important to investigate these coordinately expressed genes. A systematic, unbiased approach to identification and ranking of proteins related to IL-8's expression in tumor tissues/cells may help to define the scope of IL-8's role in tumorigenicity and important related interactions with other factors.

There are over 3 million human Expressed Sequence Tag (EST) records in GenBank (Table 1), which is still growing rapidly. EST sequences in GenBank are derived from cDNA libraries generated from a vast array of tissue types including normal, disease-state, and variously treated tissues. The number of EST clones is reported to be proportional to the abundance of cognate transcripts in the tissue

or cell type used to make the cDNA library and thus the EST distribution can provide a quantitative assessment of differential expression of a gene [17]. The expanding tissue diversity and EST coverage have increased the statistical power of EST-distribution based expression analysis. An approach using EST expression data as a binary variable (present or absent in a cDNA library) has previously identified prostate cancer-associated genes [18]. We are reporting here a novel *in silico* approach for identification of genes whose mRNA are enriched in libraries where IL-8 is represented. We focused on ESTs from cDNA libraries in which IL-8 has been sequenced at least once and analyzed the EST frequency of occurrence for all other genes in these IL-8 related cDNA libraries. Those cDNA libraries were further catalogued into tumor and non-tumor libraries, which allowed us to identify genes whose expression profile is closely related to that of IL-8 in tumor tissues.

Results

Genes co-expressed with IL-8

We found that IL-8 has been sequenced in 53 dbEST cDNA libraries (Table 1 and Additional file 1). Using these cDNA libraries (referred to as IL-8 tissue), we generated a reference database for all the genes and their expression profile relationship (measure by Z-score) with IL-8. The complete list of genes is provided in additional file 2, along with a distribution table based on contig size and Z-score (Table 2). A gene could be represented by multiple contigs which represent splice variants and sequencing errors. In this database, we provide the statistics for each contig, including the number of EST clones, cDNA libraries, EST clones in IL-8 tissue, and cDNA libraries in IL-8 tissue. First we evaluated the performance of this search using the distribution analysis results for some known genes (Table 3). The co-expression of IL-8 with IL-6, GRO- α , and GRO- β in many tissues has been well documented [5,14,19], and those genes are identified here as IL-8 related genes with an IL-8 Z-score ≥ 3 . The GRO- γ and ENA-

Table 2: Contigs identified in IL-8 tissue and IL-8-tumor tissue. Contigs are catalogued using different Z-score and total clone cutoffs. Total Clone: The total number of clones from all cDNA libraries within a contig. IL-8 contigs: Contigs identified in IL-8 tissue. IL-8-tumor contigs: Contigs identified in IL-8-tumor tissue. NC: no Z-score cutoff is applied.

	Total Clone Cutoff	Z-score Cutoff			
		NC	1	2	3
IL-8 Contigs	5	15222 [#]	3130	991	352
	10	14700	2662	755	274
	50	8063	1010	249	98
	100	3243	393	102	36
IL-8 Tumor Contigs	5	10016 [^]	1987	577	198
	10	9776	1825	520	183
	50	6677	985	281	102
	100	3020	403	115	42

[#] See Additional file 2 for detailed list. [^]See Additional file 3 for detailed list.

Table 3: EST Clone distribution for some known IL-8 related genes in the IL-8 tissue. Total Clones: The total number of clones from all cDNA libraries. IL-8 Clones: The number of clones from IL-8 tissue.

Gene	Total Clones	IL-8 Clones	IL-8 Z-score
IL-6	44	24	4.1
GRO- α	52	25	3.2
GRO- β	46	22	3
GRO- γ	8	5	3.9
ENA-78	6	4	2.8

78 [14] also show a high IL-8 Z-score, although the available EST clone number for these two genes is much smaller. Successful identification of these genes demonstrates the value of our *in silico* approach. Besides Z-score, this database can be sorted in many different ways. For example, we can adjust the stringency of correlation by using the number of IL-8 cDNA libraries as a cutoff parameter.

The significance of this study lies in the generation of a complete reference database for IL-8 co-expressed genes. We present here a list of 36 genes with high EST clone count (≥ 100) and high Z-score (≥ 3) (referred to as IL-8 genes, I-1, I-2, ..., I-36) (Table 4) to illustrate the usage of this reference database. A high EST clone count is used to ensure the statistic significance of the data mining results. IL-8 tissues represent both tumor tissues/cells (IL-8-tumor) and non-tumor tissues/cells (Table 1 and Additional file 1). To generate a tumor expression profile, the percentage of IL-8-tumor clones versus IL-8 clones (R_{tumor}) was calculated for the 36 IL-8 genes. The resulting

graph is shown in Figure 1. The R_{tumor} profile reveals a bimodal distribution. In the case of IL-8, tumor clones account for approximately 17% of all IL-8 clones ($R_{tumor} = 17\%$). Eleven IL-8 genes exhibit a R_{tumor} of greater than 80%; five of these being greater than 95% (kruppel-like factor 2 (KLF2), presenilin 1 (PS1), neural proliferation and differentiation control protein-1 (NPDC1), GW112 and claudin-3). On the other hand, eight genes were not found in any IL-8-tumor related library ($R_{tumor} = 0$).

For comparison with GRO- α and GRO- β , the ratios of IL-8 clones, GRO- α clones and GRO- β clones to total clones (R_{il8} , $R_{GRO-\alpha}$ and $R_{GRO-\beta}$) for each IL-8 gene were analyzed. Approximately half of the IL-8 genes show similar correlation with all three chemokines. As an example, GW112 exhibits high correlation with IL-8, GRO- α and - β ($R_{il8} = 69\%$, $R_{GRO-\alpha} = 71\%$, $R_{GRO-\beta} = 68\%$). On the other hand, KLF2 ($R_{il8} = 71\%$, $R_{GRO-\alpha} = 16\%$, $R_{GRO-\beta} = 2\%$) and PS1 ($R_{il8} = 65\%$, $R_{GRO-\alpha} = 5\%$, $R_{GRO-\beta} = 3\%$) show

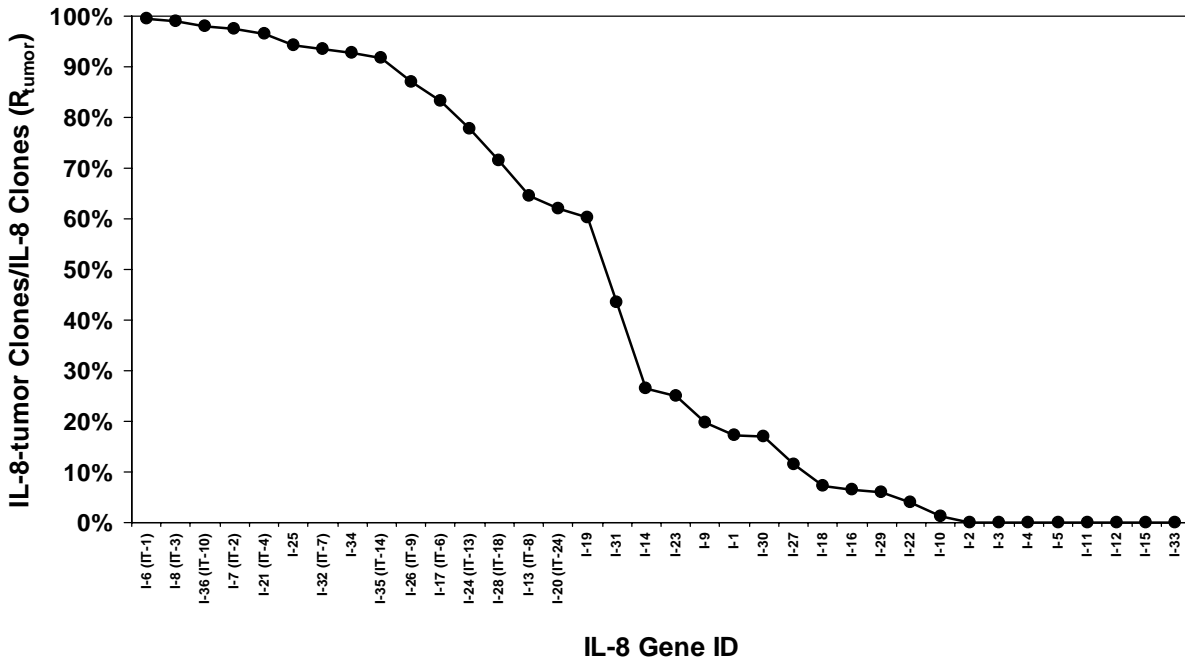


Figure 1
Distribution of IL-8 related genes in tumor cDNA libraries. The percentage ratio of IL-8-tumor clones/IL-8 clones (R_{tumor}) was calculated for each IL-8 related gene. The result is plotted here with gene id on the x-axis and R_{tumor} on the y-axis. A bimodal distribution is obtained where most genes are either highly associated (>80%) with tumor libraries or they show very little association (<10%) with tumor libraries. Five genes have an R_{tumor} value of >95%: KLF2 (99.5%), PS1 (98.9%), NPDC1 (97.9%), GW112 protein (97.4%), and claudin-3 (96.6%).

high correlation with IL-8 and only slight association with GRO- α or GRO- β .

Genes co-expressed with IL-8 in tumor tissues or cell lines

A set of tumor cDNA libraries in which IL-8 is expressed (a subset of the IL-8 tissue cDNA libraries), was generated and is referred to as IL-8-tumor tissue. This virtual tissue consists of 16 cDNA libraries (171,521 cDNA clones) (Table 1 and Additional file 1). An IL-8-tumor gene database was established based on these libraries. The complete list of genes is provided in additional file 3, along with a distribution table based on contig size and Z-score (Table 2). We present here 42 candidate genes with relatively high levels of expression in IL-8-tumor tissue (EST clone count ≥ 100 and IL-8-tumor Z-score ≥ 3) (referred to as IL-8-tumor genes, IT-1, IT-2, ..., IT-42) (Table 5). To determine

whether these genes are specifically related to IL-8-tumor tissue or whether they are also commonly found in tumors lacking IL-8 expression, the ratio between IL-8-tumor clones and general tumor clones ($R_{il8tumor}$) was plotted for IL-8-tumor genes (Figure 2). The $R_{il8tumor}$ for all IL-8 tumor genes are greater than 38%, which is much higher than the expected background ratio between IL-8-tumor clones (171521, Table 1) and general tumor clones (864428, Table 1). A few of these genes are highly specific to IL-8-tumor tissues. Most notably GW 112, PS1, and KLF2 show $R_{il8tumor}$ values of 86%, 81% and 76% respectively.

Discussion

The vast amount of available EST data allows expression analysis based solely on computational methods. It is pos-

Table 4: IL-8 related genes (IL-8 Z-score ≥ 3 and EST clone ≥ 100). Total Clones: The total number of clones from all cDNA libraries. IL-8 Clones: The number of clones from IL-8 tissue. These genes are sorted by the ratio of IL-8 clones/total clones and given a gene id (I-1, I-2, ..., I-36).

Gene ID	Definition [#]	IL-8 Clones/Total Clones (%)
I-1	IL-8 (SW:P10145)	100
I-2	serum albumin (SW:P02768)	88
I-3	fibrinogen gamma-a chain (SW:P02679)	87
I-4	aldolase B (P05062)	87
I-5	fibrinogen gamma-B chain (GI:71828)	74
I-6	kruppel-like factor 2 (SW:Q9Y5W3)	71
I-7	GW112 protein (GI:11544538)	69
I-8	presenilin 1 (SW:P49768)	65
I-9	selenoprotein P (GI:2654365)	58
I-10	beta-fibrinogen (GI:182430)	57
I-11	splice variant of serum albumin (GI:28592)	56
I-12	alpha-1-antitrypsin (SW:P01009)	56
I-13	complement factor B (GI:291922)	53
I-14	MSTP032(GI:13376832)	52
I-15	serotransferrin (SW:P02787)	51
I-16	splice variant of fibrinogen B beta (GI:14423575)	51
I-17	lumican (SW:P51884)	50
I-18	apolipoprotein A-I (SW:P02647)	50
I-19	splice variant of complement component C4A (GI:387438)	50
I-20	beta-2-microglobulin (SW:P01884)	49
I-21	claudin-3 (SW:O15551)	49
I-22	arylacetamide deacetylase (SW:P22760)	48
I-23	osteoinductive factor (SW:P20774)	48
I-24	hypothetical protein (GI:6807713)	47
I-25	tumor-associated calcium signal transducer 1 (GI:182906)	47
I-26	transgelin 2 (GI:434763)	46
I-27	complement component C4A (GI:443671)	45
I-28	secreted apoptosis related protein 1 (GI:2415415)	45
I-29	thymosin beta-4 precursor(GI:2143995)	45
I-30	secretory granule proteoglycan core protein (SW:P10124)	44
I-31	ceruloplasmin (SW:P00450)	44
I-32	nonspecific crossreacting antigen (GI:88276)	44
I-33	plasminogen (SW:P00747)	41
I-34	biglycan (GI:13279002)	41
I-35	hypothetical protein (GI:6807932)	40
I-36	neural proliferation differentiation and control protein-1 (SW:Q9NQX5)	39

[#] SW: Swissprot accession, GI: GenBank Identifier.

sible to construct a "virtual tissue" based on the expression pattern of a particular gene or group of genes (e.g. tumor marker genes). It is also possible to group genes by tissue type through combining genes from the same source tissue (e.g. tumor, brain, etc.). In this study we have generated "virtual tissues" based on (a) expression of IL-8 in all tissues and (b) expression of IL-8 in tumor tissues. Next we generated lists of genes that are most highly co-expressed with IL-8 in these two virtual tissues. Since the method is not limited to genes previously correlated with

IL-8 expression nor to known genes, we have the opportunity to identify previously overlooked correlations with known or novel genes. In addition, the relative strength of the correlations is measured by a Z-score based on the number of clones representing IL-8 or IL-8-tumor tissue compared with the total number of clones. Assuming that co-expression is related to function, these Z-scores provide a basis for ranking genes according to potential involvement in IL-8's function. This reference database for genes co-expressed with IL-8 can be cross-referenced with other

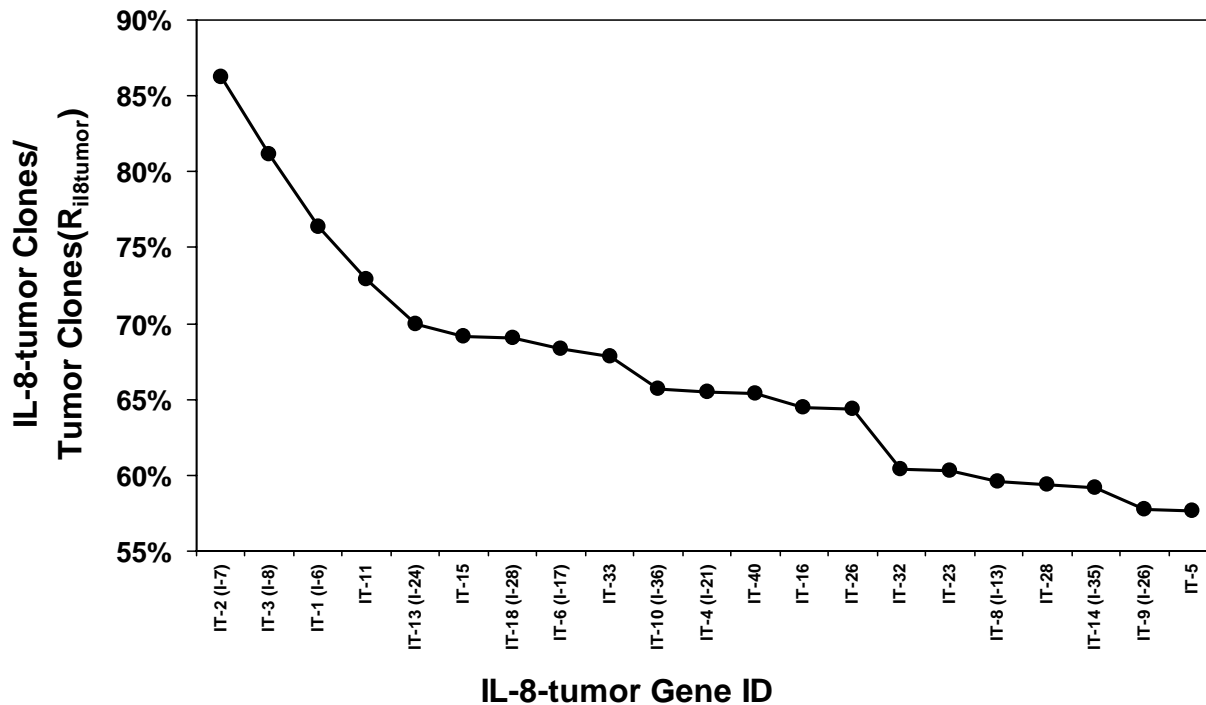


Figure 2
Distribution of IL-8-tumor related genes in all tumor cDNA libraries. The percentage ratio of IL-8-tumor clones/ tumor clones ($R_{IL8tumor}$) was calculated for IL-8-tumor related genes. The result is plotted here with gene id on the x-axis and $R_{IL8tumor}$ on the y-axis, and is ordered by decreasing $R_{IL8tumor}$. Only the top half genes (21) with relatively high $R_{IL8tumor}$ are displayed. The 3 most specific genes are GWI12 protein (86.2%), PS1 (81.1%), and KLF2 (76.4%).

large scale expression analyses, such as microarray experiments, to help decipher the regulation network and the functions of IL-8 and IL-8 co-expressed genes.

Among the genes most highly correlated with IL-8 in tumor tissues are KLF2, PS1, VEGF, and tumor necrosis factor receptor-1 (TNFR-1), lumican (keratin sulfate proteoglycan), claudin-3, and perlecan. Cytokines (IL-2 and IL-7) induce the expression of KLF2 in activated T cells which correlates with their survival [20]. They suggest that KLF2 may be involved in avoiding activation-induced cell death. It is possible that IL-8 may also mediate similar actions. This type of action could increase the ability of tumor cells to survive under conditions where cells normally apoptose.

Lumican expression has previously been correlated with higher tumor grade, lower estrogen levels in the tumor and younger age of patients in human breast cancer [21]. In addition, lumican is structurally related to both biglycan which is among the top IL-8-tumor related genes and the human embryonal carcinoma marker antigen TRA-1-60 [22]. These relationships support the hypothesis that lumican may be a tumor related protein involved with IL-8's tumorigenic function.

PS1 was also found to be highly correlated with IL-8 in both the entire list of tissues and specifically in tumor tissues. Based on EST distribution, presenilin 2 (PS2) is less related to IL-8-tumor libraries (Z-score = 0.3, IL-8-tumor clone/Total Clone = 10%). In the breast tumor microarray experiments, unlike PS1, the peak level of expression for PS2 is not in BT-549. PS1 is a gene involved in early-onset

Table 5: IL-8-tumor related genes (Zscore ≥ 3 and EST clone ≥ 100). Total Clones: The total number of clones from all cDNA libraries. IL-8 Tumor Clones: The number of clones from IL-8-tumor tissue. These are sorted by the ratio of IL-8-tumor clones/total clones and given a gene id (IT-1, IT-2, ..., IT-42).

Gene ID	Definition#	IL-8 Tumor Clones/Total Clones (%)
IT-1	kruppel-like factor 2 (SW:Q9Y5W3)	70
IT-2	gw112 protein (GI:11544538)	68
IT-3	presenilin 1 (SW:P49768)	65
IT-4	claudin-3 (SW:O15551)	47
IT-5	similar to EGP314 (GI:6678752)	43
IT-6	lumican (SW:P51884)	42
IT-7	nonspecific crossreacting antigen precursor (GI:88276)	41
IT-8	complement factor B (GI:291922)	40
IT-9	transgelin 2 (SW:P37802)	40
IT-10	neural proliferation, differentiation and control protein-1 (SW:Q9NQX5)	39
IT-11	similar to atrophin-1 (GI:1732417)	38
IT-12	immunoglobulin mu chain (GI:553361)	37
IT-13	hypothetical protein (GI:6807713)	37
IT-14	hypothetical protein (GI:6807932)	36
IT-15	alpha-1 type III collagen (GI:180414)	36
IT-16	heparan sulfate proteoglycan perlecan (GI:11602963)	35
IT-17	immunoglobulin alpha-1 heavy chain constant region (GI:184749)	35
IT-18	secreted apoptosis related protein 1 (GI:2415415)	32
IT-19	anterior gradient 2 (Xenopus laevis) homolog (GI:3779197)	31
IT-20	elongation factor 1-alpha 1 (SW:P10126)	31
IT-21	signal recognition particle 9 kd protein (SW:P49458)	31
IT-22	similar to CDK5 activator-binding protein C53 (GI:10435740)	31
IT-23	alpha-2-macroglobulin (SW:P01023)	31
IT-24	beta-2-microglobulin (GI:179318)	30
IT-25	vascular endothelial growth factor (GI:3712669)	30
IT-26	unknown	30
IT-27	similar to UNC-93 (GI:4263743)	30
IT-28	human C4 complement (GI:387438)	30
IT-29	t-complex protein 1, eta subunit (SW:Q99832)	29
IT-30	plasma protease C1 inhibitor (SW:P05155)	28
IT-31	tumor necrosis factor receptor 1 (SW:P19438)	28
IT-32	DRAL gene product (GI:1160932)	28
IT-33	unknown	28
IT-34	AF-6 (GI:3452572)	28
IT-35	ubiquitin (SW:P02248)	27
IT-36	unnamed protein product (GI:7023247)	27
IT-37	45 kda calcium-binding protein (SW:Q61112)	27
IT-38	ERF-2 (GI:509778)	27
IT-39	E74-like factor 3 (GI:1754538)	26
IT-40	connective tissue growth factor (GI:984956)	26
IT-41	mouse phospholipase D3 (GI:7242181)	25
IT-42	KIAA1536 protein (GI:7959339)	25

SW: Swissprot accession, GI: GenBank Identifier.

familial Alzheimer's disease. There is accumulating evidence that mutations in PS1 accelerate neurodegeneration and facilitate apoptosis, and some researchers [23] suggest an association with the p53 signal transduction pathway. It was suggested that down-regulation of PS1 by wildtype

p53 and also p21^{WAF-1} may be independent mechanisms leading to apoptosis and tumor suppression [24]. Our data correlating PS1 with IL-8-tumor related tissue also suggest a potential anti-apoptotic role for PS1 in tumorigenesis.

Vascular endothelial growth factor (VEGF), like IL-8, is a potent mediator of angiogenesis. VEGF has been shown to regulate angiogenesis and metastasis of bladder cancer [25] and several recent studies have reported correlations with IL-8 in several cell types including bladder cancer (TCC), non-small cell lung cancer (NSCLC) [26,27], human brain microvascular endothelial cells (HBMECS) [28] and monocytes [29]. These data collectively demonstrate the correlation of VEGF with IL-8 in tumor tissues.

TNF- α is a well-characterized, potent inducer of IL-8 transcription and an anti-apoptotic agent [30–32]. Roebuck provides a detailed analysis of IL-8 promoter structure including a promoter recruitment mechanism which involves TNF- α and the cooperativity of NF- κ B and NF-IL-6 binding sites [4,32]. TNF- α exerts many of its effects through TNFR-1 and TNFR-2 receptors, thus co-expression of these receptors with IL-8 might be anticipated. Our analysis does reveal a correlation between TNFR-1 and IL-8 expression. Using another bioinformatic approach, Eidelman et al. report a close relationship between IL-8 secretion in cystic fibrosis cells and expression of genes from the TNFR-1/NF κ B pathway. As is the case for several other genes found in this study, the TNFR-1/NF κ B pathway is associated with p53/p21^{WAF-1} tumor suppressor systems [33–37].

Our analysis identified several interesting genes co-expressed with IL-8 in tumor tissues which have not previously been associated with IL-8. Along with lumican, claudin-3, KLF2, PS1, VEGF and TNFR-1 mentioned above, these include several secreted proteins (secreted apoptosis related protein 1 and NPDC1) which are also likely to function in coordination with IL-8; several genes from the well-known family of complement factors (C1 esterase, human complement C4A and complement factor B); and several unknown or hypothetical genes. Interestingly, several of the genes correlate with the p53/p21^{WAF-1} tumor suppressor network. We expect that some of these genes may present important therapeutic targets at one or more levels in the network of IL-8 mediated interactions related to tumorigenesis.

Like all EST-based expression analysis, there are limitations for this *in silico* method. First, EST-based expression analysis is not suitable for transcripts with low abundance (no or few representative ESTs in the database), which precludes statistical analysis. Second, EST-based expression analysis can only indicate the co-expression of genes, but can not accurately measure the expression level. Third, to generate a comprehensive reference database, we decided not to exclude cDNA libraries based on the sequence depth or special manipulation (e.g. subtraction and normalization). The depth of sequencing and special manip-

ulation are not consistent across cDNA libraries, which could also influence the sensitivity of the analysis.

The computational expression analysis methods developed here are not limited to the identification of IL-8 related genes, but can also be applied to many other proteins of interest. This method is complementary to other large-scale expression analysis methods (e.g. microarray) in that it is not limited by the physical presence of a gene on microarray, thus it offers a unique approach to discovering potential functional links between genes through expression profiling.

Methods

In silico expression analysis

All non-commercial software used in these studies was written in PERL 5.0. Human EST sequence and cDNA library information were retrieved from GenBank (Release 120) and an in-house relational database model (Sybase, SQL Server Release 11.0, CA, Sybase Inc.) was created to mirror the public human EST database (dbEST). The EST sequences were first binned into clusters if they share the same 21 mer tag beginning with ATG or CTG. Next, ESTs in each cluster were assembled into contigs using the PHRAP sequence assembly software (Phil Green, Unpublished). Based on the ESTs corresponding to IL-8 in its cognate contig, we generated a set of cDNA libraries in which IL-8 is represented. This set of libraries is referred to as IL-8 tissue and can be considered as a virtual tissue consisting of 53 cDNA libraries (306,888 cDNA clones) (Table 1 and Additional file 1). We chose contigs consisting of at least 5 EST clones and present in at least 3 IL-8 libraries. The frequency of occurrence (F) and Z-score for all the contigs in these cDNA libraries were calculated as:

$$F = \frac{N_{IL8}}{N_{total}}$$

$$Z\text{-score} = \frac{F - \sum F}{SDV}$$

where N_{IL8} is the number of EST clones in IL-8 tissue for a particular contig, N_{total} is the total EST clone count in that contig. Contigs were collected into groups based on contig size (number of EST clones in a contig), and small contig groups (< 1000 contigs) were merged together with neighboring groups (similar in contig size) to make sure that there were at least 1000 contigs in each group. The rationale for grouping contigs based on size is that the distribution pattern of F (i.e. mean and Z-score) may vary for

contigs of different size. n is the total number of contigs in a contig group, ΣF is the sum of F in a contig group, and SDV is the standard deviation of F in a contig group. Assuming a Gaussian Distribution, the fractions of the population that are greater than Z -score SDV and also above the mean are 15.87%, 2.28%, and 0.13% for Z -scores of 1, 2, and 3 respectively. Similarly, a set of tumor cDNA libraries (based on their GenBank Annotations) in which IL-8 is represented was generated. This set of libraries is referred to as IL-8-tumor tissue, consisting of 16 cDNA libraries (171,521 cDNA clones) (Table 1 and Additional file 1). Using the method described above, a database for IL-8-tumor candidate genes (using IL-8-tumor tissue) was generated and IL-8-tumor Z -scores were calculated. Databases for GRO- α and GRO- β related genes were generated in a similar way using the cDNA libraries where GRO- α or GRO- β were represented respectively (Table 1).

Authors' contributions

LB and LUW carried out the in silico expression analysis and drafted the manuscript. SL participated in the cataloging of tumor EST libraries. LYW, RB, JG, JH, PQ, EG, ML, and MK participated in design of the study. All authors read and approved the final manuscript.

Additional material

Additional File 1

IL-8 cDNA Libraries List of all cDNA libraries in which IL-8 has been sequenced at least once.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-3-29-S1.doc>]

Additional File 2

Contigs identified in IL8 tissue List of contigs that have at least one EST in IL8 tissue.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-3-29-S2.xls>]

Additional File 3

Contigs identified in IL8-tumor tissue List of contigs that have at least one EST in IL8-tumor tissue

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-3-29-S3.xls>]

References

- Moore BB, Arenberg DA, Addison CL, Keane MP, Strieter RM: **Tumor angiogenesis is regulated by CXC chemokines.** *J Lab Clin Med* 1998, **132**:97-103
- Arenberg DA, Polverini PJ, Kunkel SL, Shanafelt A, Hesselgesser J, Horuk R, Strieter RM: **The role of CXC chemokines in the regulation of angiogenesis in non-small cell cancer.** *J Leukoc Biol* 1997, **62**:554-562
- Shi Q, Abbruzzese JL, Huang S, Fidler IJ, Xiong Q, Xie K: **Constitutive and Inducible Interleukin 8 Expression by Hypoxia and Acidosis Renders Human Pancreatic Cancer Cells More Tumorigenic and Metastatic.** *Clin Cancer Res* 1999, **5**:3711-3721
- Roebuck KA: **Regulation of Interleukin-8 Gene Expression.** *J Interferon and Cytokine Res* 1999, **19**:429-438
- Bar-Eli M: **Role of Interleukin 8 in Tumor Growth and Metastasis of Human Melanoma.** *Pathobiology* 1999, **67**:12-18
- Gaffy G, Mohammed KA, Dowling PA, Nasreen N, Ward MJ, Antony VB: **Interleukin 8: An Autocrine Growth Factor for Malignant Mesothelioma.** *Cancer Res* 1999, **59**:367-371
- Arenberg DA, Kunkel SL, Polverini PJ, Glass M, Burdick MD, Strieter RM: **Inhibition of Interleukin-8 Reduces Tumorigenesis of Human Non-Small Cell Lung Cancer in SCID Mice.** *J Clin Invest* 1996, **97**:2792-2802
- Inoue K, Slaton JW, Kim SJ, Perrotte P, Eve BY, Bar-Eli M, Radinsky R, Dinney CPN: **Interleukin 8 Expression Regulates Tumorigenicity and Metastasis in Human Bladder Cancer.** *Cancer Res* 2000, **60**:2290-2299
- Kupper TS, Ballard D, Chua AO, McGuire JS, Flood P, Gubler U: **Human keratinocytes contain mRNA indistinguishable from monocyte interleukin 1 mRNA.** *J Exp Med* 1996, **164**:2095-2100
- Oxholm A, Oxholm P, Staberg B, Bendtsen K: **Immunohistochemical detection of interleukin-1-like molecules and tumor necrosis factor in human epidermis before and after UV-B irradiation in vivo.** *Br J Dermatol* 1988, **118**:369-376
- Kirnbaur R, Kock A, Schwartz T, Urbanski A, Krutman J, Borth Q, Ansel JC, Luger TA: **Interferon- β 2, B-cell differentiation factor 2, hybridoma growth factor (interleukin-6) is expressed and released by human epidermal cells and epidermoid carcinoma cell lines.** *J Immunol* 1989, **142**:1922-1928
- Yarr M, Palleroni AV, Gilchrist BA: **Keratinocytes contain an interferon like protein that may modulate their growth and differentiation.** *Ann N Y Acad Sci* 1988, **548**:299-311
- Shih IM, Elder DE, Hsu MY, Herlyn M: **Regulation of Mel-CAM/MUC18 Expression on Melanocytes of Different Stages of Tumor Progression by Normal Keratinocytes.** *Am J Pathol* 1994, **145**:837-845
- Walz A, Burgener R, Car B, Baggiolini M, Kunkel SL, Strieter RM: **Structure and Neutrophil-activating Properties of a Novel Inflammatory Peptide (ENA-78) with Homology to Interleukin 8.** *J Exp Med* 1991, **174**:1355-1362
- Haskill S, Peace A, Morris J, Sporn SA, Anthony A, Lee SW, Smith T, Martin G, Ralph P, Sager R: **Identification of three related human GRO genes encoding cytokine functions.** *Proc Natl Acad Sci* 1990, **87**:7732-7736
- Belperio JA, Kean MP, Arenberg DA, Addison CL, Ehlert JE, Burdick MD, Strieter RM: **CXC chemokines in angiogenesis.** *J Leukoc Biol* 2000, **68**:1-8
- Audic S, Claverie J: **The significance of Digital Gene Expression Profiles.** *Genome Res* 1997, **7**:986-995
- Walker MG, Volkmuth W, Sprinzak E, Hodgson D, Klinger T: **Prediction of Gene Function by Genome-Scale Expression Analysis: Prostate Cancer-Associated Genes.** *Genome Res* 1999, **9**:1198-1203
- Loukinova E, Dong G, Enamorado-Ayala I, Thomas GR, Chen Z, Schreiber H, Van Waes C: **Growth Regulated Oncogene- α expression by murine squamous cell carcinoma promotes tumor growth, metastasis, leukocyte infiltration and angiogenesis by a host CXC receptor-2 dependent mechanism.** *Oncogene* 2000, **19**:3477-3486
- Schober SL, Kuo CT, Schluns KS, Lefrancois L, Leiden JM, Jameson SC: **Expression of the Transcription Factor Lung Kruppel-Like Factor is Regulated by Cytokines and Correlates with Survival of Memory T Cells In Vitro and In Vivo.** *J Immunol* 1999, **163**:3662-3667
- Leygue E, Snell L, Dotzlaw H, Hole K, Hiller-Hitchcock T, Roughley PJ: **Expression of Lumican in Human Breast Carcinoma.** *Cancer Res* 1998, **58**:1348-1352
- Badcock G, Pigott C, Goepel J, Andrews PW: **The human embryonal carcinoma marker antigen TRA-1-60 is a sialated keratan sulfate proteoglycan.** *Cancer Res* 1999, **59**:4715-4719
- Amson R, Lasalle JM, Halley H, Prieur S, Lethrosne F, Roperch JP, Israëli D, Gendron MC, Duyckaerts C, Checler F, Dausset J, Cohen D, Oren M, Teclerman A: **Behavioral alterations associated with**

- apoptosis and down-regulation of presenilin 1 in the brains of p53-deficient mice.** *Proc Natl Acad Sci* 2000, **97**:5346-5350
24. Roperch JP, Alvaro V, Prieur S, Tuynder M, Nemani M, Lethrosne F, Piouffre L, Gendron MC, Israeli D, Dausset J, Oren M, Amson R, Tellerman A: **Inhibition of presenilin 1 expression is promoted by p53 and p21^{WAF-1} and results in apoptosis and tumor suppression.** *Nat Med* 1998, **4**:835-838
 25. Inoue K, Slaton JW, Davis DW, Hicklin DJ, McConkey DJ, Karashima T, Radinsky R, Dinney CPN: **Treatment of Human Metastatic Transitional Cell Carcinoma of the Bladder in a Murine Model with the Anti-Vascular Endothelial Growth Factor Receptor Monoclonal Antibody DC101 and Paclitaxel.** *Clinical Cancer Res* 2000, **6**:2635-2643
 26. Masuya D, Huang C, Liu D, Kameyama K, Hayashi E, Yamauchi A, Kobayashi S, Haba R, Yokomise H: **The intratumoral expression of vascular endothelial growth factor and interleukin-8 associated with angiogenesis in nonsmall cell lung carcinoma patients.** *Cancer* 2001, **92**:2628-2638
 27. Yuan A, Yu CJ, Luh KT, Kuo SH, Lee YC, Yang PC: **Aberrant p53 Expression Correlates With Expression of Vascular Endothelial Growth Factor mRNA and Interleukin-8 mRNA and Neoangiogenesis in Non-Small-Cell Lung Cancer.** *J Clinical Oncol* 2002, **20**:900-910
 28. Lee TH, Avraham H, Lee SH, Avraham S: **Vascular endothelial growth factor modulates neutrophil transendothelial migration via upregulation of Interleukin-8 in human brain microvascular endothelial cells.** *J. Biol. Chem.* 2002, **277**:10445-10451
 29. Tezono K, Sarker KP, Kikuchi H, Nasu M, Kitajama I, Maruyama I: **Bioactivity of the vascular endothelial growth factor trapped in fibrin clots: production of IL-6 and IL-8 monocytes by fibrin clots.** *Haemostasis* 2001, **31**:71-79
 30. Singh RK, Varney ML: **Regulation of Interleukin 8 Expression in Human Malignant Melanoma Cells.** *Cancer Res* 1998, **58**:1532-1537
 31. Nagaki M, Sigiyama A, Osawa Y, Naiki T, Nakashimi S, Nozawa Y, Moriwaki H: **Lethal hepatic apoptosis mediated by tumor necrosis factor receptor, unlike Fas-mediated apoptosis, requires hepatocyte sensitization in mice.** *J Hepatol* 1999, **31**:997-1005
 32. Matsusaka T, Fujikawa K, Nishio Y, Mukaida N, Matsushima K, Kishimoto T, Akira S: **Transcription factors NF-IL-6 and NF-kappa B synergistically activate transcription of the inflammatory cytokines, interleukin 6 and interleukin 8.** *Proc Natl Acad Sci USA* 1993, **90**:10193-10197
 33. Saile B, Matthes N, El Armouche H, Neubauer K, Ramadori G: **The bcl, NFkappaB and p53/p21WAF1 systems are involved in spontaneous apoptosis and in the anti-apoptotic effect of TGF-beta or TNF-alpha on activated hepatic stellate cells.** *Eur J Cell Biol* 2001, **80**:554-561
 34. Rokhlin OV, Gudov AV, Kwek S, Glover RA, Gewies AS, Cohen MB: **p53 is involved in tumor necrosis factor-alpha-induced apoptosis in the human prostatic carcinoma cell line LNCaP.** *Oncogene* 2000, **19**:1959-1968
 35. Donato NJ, Perez M: **Tumor Necrosis Factor-induced Apoptosis Stimulates p53 Accumulation and p21WAF1 Proteolysis in ME-180 Cells.** *J Biol Chem* 1998, **273**:5067-5072
 36. Gotlieb WH, Watson JM, Rezai A, Johnson M, Martinez-Maza O, Berrek JS: **Cytokine-induced modulation of tumor suppressor gene expression in ovarian cancer cells: Up-regulation of p53 gene expression and induction of apoptosis by tumor necrosis factor-alpha.** *Am J Obstet and Gynecol* 1994, **170**:1121-1130
 37. Jeoung DI, Tang B, Sonenberg M: **Effects of Tumor Necrosis Factor-alpha on Antimitogenicity and Cell Cycle-related Proteins in MCF-7 Cells.** *J Biol Chem* 1995, **270**:18367-18373

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

editorial@biomedcentral.com