

Analysis of Regulatory Genomics and Gene Expression Pattern of Medicinal Importance Genes of *Helicobacter Pylori*

Pramod Katara*, Ashutosh Singh, Dhvani Ragav, and Vinay Sharma

Department of Bioscience and Biotechnology, Banasthali University, Bansthali 304022, India

Abstract

Helicobacter pylori has is recognized as the main causal agent of chronic gastritis and duodenal ulcers, and it is associated with the subsequent development of gastric carcinoma. It adapted to life in a unique niche, the gastric epithelium of primates, its promoter may there for show different types of regulatory motif from those of other bacteria and it is well known fact that motif are the sequence portion which are responsible for gene regulation, by studying them we can control the expression of such genes of interest.

Here, the objective of this work is to analyze the regulatory sequence pattern of virulence genes that have medicinal importance for providing a basis for drug development process and further analysis of transcriptional regulatory networks.

For this purpose using available microarray gene expression data from Stanford Microarray Database, and computation tools, As a result we found that *helicobacter pylori* shows different type of regulatory motif of *Oligo* and *Dyad* pattern in studied genes. The most common length of single block, *Oligo* motif is 8 – 14, and the most common pattern for *Dyad* is 4 (4/8)3, we also observe that the GC content of these regulatory is just 15-20%, which is comparatively very less.

Keywords: Transcription factor binding sites; Open Reading Frame; Gene expression pattern; Co-expressed gene

Introduction

In recent year there has been an exponential growth in molecular genetics technologies, such as DNA microarrays, allow us for the first time to obtain a “Whole” view of the cell. At the same time development in high throughput computing and bioinformatics provide various computational algorithm and tool to define functional, structural and regulatory behavior of such genes which are analyzed by DNA microarray technology.

At present there are various different technologies for measuring gene (mRNA) expression levels are present, but cDNA microarrays are preferably used in scientific community, this allows for a quantitative readout of gene expression on a gene-by-gene basis (Brown and Botstein, 1999; Duggan et al., 1999). Microarrays have opened the possibility of creating the possibility of creating data sets of molecular information to represent many systems of biological or clinical interest. Gene expression profiles can be used as input to large-scale data analysis such as – to discover regulatory genomics, to discover taxonomy, to discover new gene of drug importance, and to increase our understanding of normal and disease states (Debouck and Goodfellow, 1999; Alizadeh et al., 2000).

A first step to analyze this all type of information is to examine the extremes, i.e. genes with significant differential expression in two individual samples or in a time series after a given treatment. This simple technique can be extremely efficient, for example, in screens for potential tumor markers or drug targets (Debouck and Goodfellow, 1999). However, such analyses do not address the full potential of genome-scale experiments to alter our understanding of cellular biology by providing, through an inclusive analysis of the entire repertoire of transcripts, a continuing comprehensive window into the state of a cell as it goes through a biological process. What is needed instead is a holistic approach to analysis of genomic data that focuses on illuminating order in the entire set of observations, allowing biologists to develop an integrated understanding of the process being studied.

A natural basis for organizing gene expression data is to Clustering genes with similar patterns of expression (Eisen et al., 1998; Alizadeh et al., 2000). The first step for it is to adopt a mathematical description of similarity. For any series of measurements, a number of sensible measures of similarity in the behavior of two genes can be used, such as the Euclidean distance, angle, or dot products of the two n -dimensional vectors representing a series of n measurements. We have found that the standard correlation coefficient (i.e., the dot product of two normalized vectors) conforms well to the intuitive biological notion of what it means for two genes to be “co-expressed;” this may be because this statistic captures similarity in “shape” but places no emphasis on the magnitude of the two series of measurements.

Conserved DNA sequences are present in all types of organism as Motif or in any other form, which correspond to transcriptional regulatory motifs in upstream regions of genes (McGurie et al., 2001). These conserved regions are often binding sites for DNA-binding proteins and some time also work as TFBS for transcription factor and known as gene regulatory elements.

In bacteria it is difficult to locate the regulatory region for a gene found within an operon (Jacob and Monod, 1961), since the promoter for that operon can lay several genes upstream, and it is difficult to predict which gene is at the head of the operon

*Corresponding author: Pramod Katara, Department of Bioscience and Biotechnology, Banasthali University, Bansthali 304022, India, Tel: +919413094705; E-mail: pmkatara@gmail.com

Received November 05, 2009; Accepted February 06, 2010; Published February 06, 2010

Citation: Katara P, Singh A, Ragav D, Sharma V (2010) Analysis of Regulatory Genomics and Gene Expression Pattern of Medicinal Importance Genes of *Helicobacter Pylori*. J Comput Sci Syst Biol 3: 010-015. doi:10.4172/jcsb.1000049

Copyright: © 2010 Katara P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(Price et al., 2005). In addition, there are fewer instances of most regulatory motifs in a bacterial genome than in the *yeast* or any other Eukaryotes, as there is usually only one instance of a regulatory motif per operon instead of one instance per gene. It is easier to discover a motif that is found in more copies in the genome. However, one can increase the number of instances of a conserved regulatory motif by pooling together upstream sequence from co-express (co-regulated) or orthologous genes in closely related organisms, assuming the motif is conserved across these organisms.

Now days in this bioinformatics era, various statistical based software are available which are based on different algorithm, for analysis of such conserved motif in given upstream sequences which work as regulatory elements, but main thing is to get biological relevance from them (Lescot et al., 2002; McGurie et al., 2001; Hughes et al., 2000; Hu et al., 2004; Fogel et al., 2004; Fogel et al., 2008).

Considering the above assumption, we have designed an analysis of gene expression pattern and regulatory genomics of some medicinal important genes of *helicobacter pylori* (small gram-negative bacteria), which is a pathogenic bacteria and infect over 50% of the global population by causing disease like ulcer (Matysiak-Budnik and Megraud, 1994) and gastric cancer (Graham and Yamaoka, 1999; Aguilar et al., 2001; Parkin, 2001; Ohata et al., 2004). Here we try to find out the consensus nucleotide elements (motifs) and there pattern, which are responsible for expression of gene and work as regulatory sequence in *helicobacter pylori*.

Helicobacter pylori is a very significant organism for medical purpose, the whole genome sequence of its widely available strain (26695 and J99) are completely sequenced, they contain a single circular genome of 1.7 million base pairs and around 1,500 predicted coding sequences (Tomb et al., 1997; Alm et al., 1999). The objective of present work is to find gene expression pattern of virulence genes and their regulatory sequence which work as binding sites for Transcriptional regulators in virulent genes. This study is based on well established concept that genes with similar gene expression patterns are most probably share common regulatory machinery (Altman and Raychaudhuri, 2001; Allocco et al., 2004).

Results and Discussion

Prediction of coexpressed genes

Here we have obtained a Hierarchical cluster as output by using microarray gene expression data in cluster which can visualize in tree-view as a hierarchical tree. We have found that this tree contain all given data in a hierarchical form. According to gene expression value, closely related (co-express) gene would in same cluster. By using different correlation type we also found that the centered correlation is better and suitable for hierarchical clustering and gives more appropriate output for further process (Eisen et al., 1998; Wen et al., 1998).

After clustering, we observed genes of all cluster and found that most of the well documented virulence genes i.e. *VacA*, *FecA* and some *Cag-PAI*, are shows proximity and most of them are grouped in two cluster along with some other genes, so these two cluster are of our interest. The genes of these selected clusters 'seed cluster' used for further analysis and rest of the cluster

were discarded. This both cluster contains different number of gene which shows same gene expression.

The numbers of gene par cluster are as follows

CL1 - 17 genes

HP1458; HP0547; HP1588; HP0010; HP0109; HP0111; HP0110; HP0631; HP024 ; HP065; HP0633; HP0632; HP1006; HP0601; HP0389 ; HP0390 ; HP0673

CL 2- 19 genes

HP0382; HP0144; HP1400; HP0668; HP1178; HP0920; HP1038; HP1037; HP1342; HP0227; HP0448; HP1502; HP0065; HP0527; HP0887;HP0722; HP1512; HP0541; HP1177

Analysis of gene expression pattern of coexpressed gene

By applying clustered gene in genesis and some other plotting option for gene expression pattern analysis, plot by them shows that according to time period and provided environment condition, the expression of gene become changed, and this change is observable. This fluctuation seems same for most of genes which are is same cluster (Cluster 1 and 2) Figure 1.

Gene expression pattern shows that, when we going to calculate the variance for all gene at every given condition, we found that at some point it is very high for some gene. It is shows that after clustering there is a chance of getting some false positive gene in cluster. For those genes whose function is tilled not known but are came in these two cluster, we can assume that they are somehow related with the virulence of *H pylori*, and consider as probable virulence genes (Eisen et al., 1998; Spellman et al., 1998; van-Noort et al., 2003).

Retrieval of upstream (promoter) region

By help of RSAT we retrieve the promoter region for every cluster; here we found that the length of upstream region is differing from gene to gene. But the considerable thing which we observe here is that, for some genes, the length of promoter region is '0' (CL 1- 5; CL 2- 3), and they shows the distance '0' with there neighbor, when we search for it in TIGR Operon Database, found that most of such gene are related to same Operon unit (Salgado et al., 2000). Its give indication that probably ORF which shows '0' distance from there neighbor, are controlled by the same regulatory machinery, and by present method we can't be able to find motif for them because we didn't get Promoter region for them.

Analysis of Regulatory Motif

As a result of oligo analysis (regulatory motif) using MEME which use Position weight matrices, we predict five most probable motif for all genes in both cluster. In some cases it may be possible that, predicted sequence may be 'Pribnow box - TATAAT' (Pribnow, 1975), which is comman and essential to start transcripton of all genes, Thus we need attention when consider this type of sequence as a regulatory motif. By considering motif score we make merit of probable motif (Table 1). It is hard to say that, which particular size of motif is more probable in cluster, but here in MEME this is based on P-Value and E-Value (Bailey et al., 2006).

For confirmation we also find out the regulatory motif by

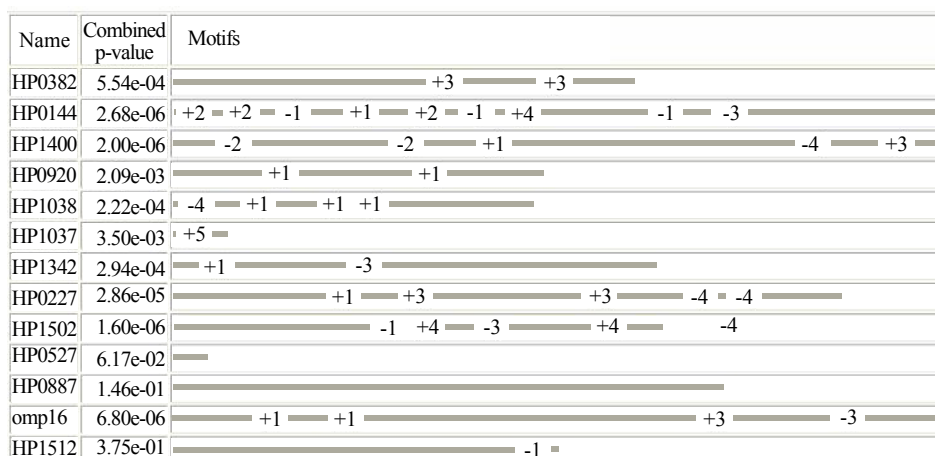


Figure 1a: It shows various motifs in form of number along with their direction and position in cluster 1.

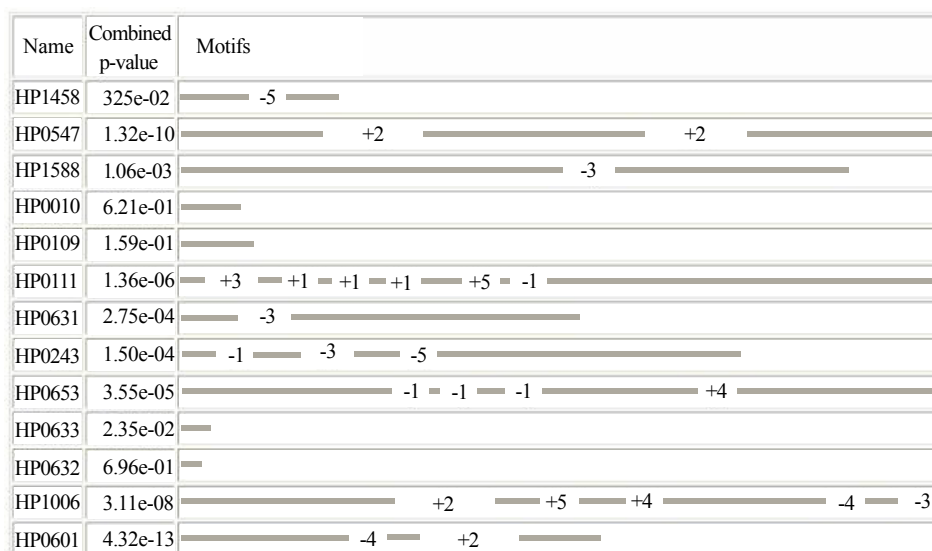


Figure 1b: It shows various motifs in form of number along with their direction and position in cluster 2.

Cluster-1 MOTIFS (nucleotide)	Cluster-2 MOTIFS (nucleotide)
MOTIF / WIDTH / BEST POSSIBLE MATCH	MOTIF / WIDTH / BEST POSSIBLE MATCH
1 11 TTTTAGGGGTA	1 6 AAGGGG
2 11 GCGCCGGCCCC	2 40 ACTTGAATCGACCATGCTTCCAACCCACTA CAGCGATGC
3 12 TTTTATGGTAAA	3 21 CATTTGCTTAAAAAGCCGTTT
4 8 TCCCATGA	4 11 GCCGGTGTCT
5 6 AGGGGG	5 6 ACCCCA

Table 1: It shows all five motifs of cluster-1 and 2 along with their width and nucleotide sequence.

Bioprosppector also and observe that more than 60% motif from MEME are matched with this Bioprosppector motif (Liu et al., 2001). With MEME considerable thing is that, it may give different result with different run because of its learning capability. Thus by applying Bioprosppector we confirm the presence of such motifs.

The motifs in genes of both cluster are present in different numbers and pattern, in some genes they did not show their presence even a single time i.e. in cluster-1 HP0527 and HP0887, and in cluster-2 HP0010, HP0109, HP0633 and HP0632, and in some other genes they are present in repeated manner i.e. in cluster-1 gene HP0382 shows two time presence of motif '+3' and in cluster-2 HP0111 shows three time presence of motif '+1' continuously, here we also observe that motifs are present in both

'+' and '-' directions (Figure 1a and 1b). This all shows that coexpressed genes not share all of these motifs with each other, they share only some specific motifs in their upstream region and only because of them they are coregulated, or their may be some other factors behind their coregulation (Kremling et al., 2000).

Analysis of dyad pattern motif

Here by help of RSAT (dyad analysis option) we found some possible Dyad pattern motif, but the number of such dyad is very less in both clusters (Table 2 and Table 3).

By observation of these tables found that, If we consider only non overlapping dyad then it is just 3 (CL 1 =1, CL 2 = 2) out of all 36 genes. The commonly visited pattern of Dyad is 4 (4/8)3.

Dyad type	dyad_sequence	occ_P	occ_E	occ_sig	Overlapping	remark
Any repeat	aaan{8}ttd	5.7e-06	1.2e-02	1.94	allow	inv_rep
Any repeat	aaan{9}ttd	3.2e-05	6.5e-02	1.19	Prevent	-
Any repeat	aaan{7}ttd	0.00031	6.3e-01	0.20	allow	-
Any dyad	aaan{8}ttd	5.7e-06	2.5e-01	0.60	allow	inv_rep

Table 2: Identified dyad sequences of Cluster-1 along with there occurrence probability, occurrence significance and other properties.

Dyad type	dyad_sequene	occ_P	occ_E	Occ_sig	Overlapping	remark
Any repeat	ctcn{1}ctc	2.6e-05	5.2e-02	1.28	Prevent	dir_rep
Any repeat	ctcn{3}ctc	0.00012	2.4e-01	0.63	allow	dir_rep
Any repeat	ctcn{5}ctc	0.00015	3.0e-01	0.52	Prevent	dir_rep
Any dyad	agan{2}gag	1.4e-06	6.0e-02	1.22	allow	-

Table 3: Identified dyad sequences of Cluster-2 along with there occurrence probability, occurrence significance and other properties.

The range of occurrence significance is from 0.01 to 2.80, Indicated that the significance of most of observed dyad is not too much, thus we can conclude that most of the virulence genes of *H pylori* are regulated by simple oligo pattern regulatory motifs and dyad pattern motifs not shown any significant role in their regulation.

Conclusion

In this work, we found out the gene expression pattern, regulatory elements (motifs) of different length and pattern in upstream regions of *VacA* and *CagA*-PAI family genes along with some other genes, which are clustered together according to there gene expression value in response of different experimental conditions, here we also found out the function of some unknown genes HP1458, HP0673, HP0448 and HP0722.

The study of gene expression pattern shows that virulence genes are shows under-expression in initial adverse conditions i.e. iron but after some time they become normal/over-express, it shows that virulence genes have capability to express even in adverse condition.

By study of Regulatory genomics we found that virulence gene of *Helicobacter pylori* show a unique regulatory genomics. They shared regulatory motif of various patterns and length (6-22), the most common length of single block motif is 8 to 14 nucleotide and the most common pattern for Dyad is [4 (4/8)3]. The difference between most of found motif is just at a single or double nucleotide level. In all motifs more than 80% nucleotide shared by A & T base and the ratio of G and C base is just 15 to 20% which is less than its whole genome GC ratio (39%). The unknown gene HP0887 shows enough similarity with *VacA* gene which produces cytotoxin, indicate that somewhere HP0887 is also related with cytotoxin activity thus it had medicinal importance, Gene HP1458 shows similar gene expression pattern with *Cag26* gene which is a member of *Cag*-PAI and work as marker for the presence of PAI, thus here by this study gene HP1458 can also be used as marker for the presence of PAI.

Methodology

In the present study for the mining of the regulatory sequence of virulence genes of *Helicobacter pylori*, virulence genes are collected from various available database VFDB (Chen et al., 2005; Yoon et al., 2005), and by data mining from various research articles (Censini et al., 1996; Maeda et al., 1999; Karhukorpi et al., 2000; Ali et al., 2005).

For the analysis, data sets from the asynchrony experiments

reported on at Stanford University Web Site (<http://genome-www.stanford.edu>), is downloaded. The data is collected as fluorescence measurements which is made relative in the levels of expression in iron starvation. This data set was already used to determine genes that are regulated by iron availability and to access the growth phase dependence on that regulation (Merrell et al., 2003). On SMD this raw data are provided in 32 excel file sets (each file have equivalent weight) and each file conation expression data of different time and different iron concentration. It consists of a set of 4607 gene expression data. The raw data files are sorted and scaled by taking logarithm at base 2 of R/G normalized (mean) ratio. The genes having ratio as an absolute value greater than two standard deviation from the mean in at least one array are taken for further filtering (Tamayo et al., 1999; Ulm et al., 2004). In this sequel, only 80% of the sorted genes are selected as they pass through spot filter criteria. Ultimately a set of 772 out of 4607 genes are taken for current analysis.

For the clustering of genes on the basis of expression values (Schena et al., 1996; Yeung et al., 2004), Cluster software from Eisen lab is used (Eisen et al., 1998). As we are interested in genes responsible for virulence, therefore after hierarchical clustering, only those clusters having our gene of interest (known virulence genes) were selected and all other genes shared by such cluster were considered as probable virulence genes (van Noort et al., 2003; Bergmann et al., 2004).

Further for the analysis of the pattern of regulatory sequence of coexpressed virulence genes by using RSAT program, first we mine their upstream region, then oligo-analysis by the help of MEME and dyad-analysis using RSAT server was performed (van-Helden et al., 2000; van-Helden et al., 2003; Thomas-Chollier et al., 2008; Defrance et al., 2008), which took above obtained clusters information as an input for finding upstream region, this process is based on comparative analysis (Lenhard and Wasserman, 2002).

Acknowledgements

Authors acknowledge the bioinformatics facility at Department of Bioscience and Biotechnology, Banasthali University, Banasthali for providing essential facilities for completion of this research work.

References

- Aguilar GR, Ayala G, Fierros-Zarate G (2001) *Helicobacter pylori* recent advances in the study of its pathogenicity and prevention. Salud Publica Mex 43: 237-47. » CrossRef » PubMed » Google Scholar

2. Ali M, Khan AA, Tiwari SK, Ahmed N, Rao LV, et al. (2005) Association between cag-pathogenicity Island in *Helicobacter pylori* isolates from peptic ulcer, gastric carcinoma, and non-ulcer dyspepsia subjects with histological changes. *World J Gastroenterol* 11: 6815-22. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
3. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503-511. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
4. Allocco DJ, Kohane IS, Butte AJ (2004) Quantifying the relationship between co-expression, coregulation and gene function. *BMC Bioinformatics* 5: 18. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
5. Alm RA, Ling LS, Moir DT, King BL, Brown ED, et al. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397: 176-80. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
6. Altman RB, Raychaudhuri S (2001) Whole-genome expression analysis: challenges beyond clustering. *Curr Opin Struct Biol* 11: 340-7. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
7. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34: W369-W373. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
8. Bergmann S, Ihmels J, Barkai N (2004) Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol* 2, E9. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
9. Brown, PO, Botstein, D (1999) Exploring the new world of the genome with microarrays. *Nat Genet* 21: 33-7. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
10. Bulyk ML, McGuire AM, Masuda N, Church GM (2004) A Motif Co-Occurrence Approach for Genome-Wide Prediction of Transcription-Factor-Binding Sites in *Escherichia coli*. *Genome Res* 14: 201-208. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
11. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, et al. (1996) cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci* 93: 14648-53. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
12. Chen L, Jian Y, Jun Y, Zhijian Y, Lilian S, et al. (2005) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 33: D325-D328. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
13. Debouck C, Goodfellow PN (1999) DNA microarrays in drug discovery and development. *Nat Genet* 21: 48-50. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
14. Defrance M, Janky R, Sand O, van Helden J (2008) Using RSAT oligo-analysis and dyad-analysis tools to discover regulatory signals in nucleic sequences. *Nat Protoc* 3: 1589-1603. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
15. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM (1999) Expression profiling using cDNA microarrays. *Nat Genet* 21: 10-4. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
16. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* 95: 14863-8. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
17. Fogel GB, Porto VW, Varga G, Dow ER, Craven AM, et al. (2008) Evolutionary computation for discovery of composite transcription factor binding sites. *Nucleic Acids Res* 36: e142. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
18. Fogel GB, Weekes DG, Varga G, Dow ER, Harlow HB, et al. (2004) Discovery of sequence motifs related to coexpression of genes using evolutionary computation. *Nucleic Acids Res* 32: 3826-3835. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
19. Graham DY, Yamaoka YH (1999) Pylori and cagA: relationships with gastric cancer, duodenal ulcer, and reflux esophagitis and its complications. *Gut* 44: 336-41. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
20. Hu Z, Fu Y, Halees AS, Kielbasa SM, Weng Z (2004) SeqVISTA: a new module of integrated computational tools for studying transcriptional regulation. *Nucleic Acids Res* 32: W235-W241. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
21. Hughes JD, Estep, Preston W, Church GM (2000) Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J Mol Biol* 296: 1205-14. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
22. Jacob F, Monod J (1961) Genetic regulatory mechanism in the synthesis of proteins. *J Mol Biol* 3: 318-56. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
23. Karhukorpi J, Yan Y, Kolho KL, Rautelin H, Lahti M, et al. (2000) cagA, vacA, iceA virulence genes of *Helicobacter pylori* isolates of children in Finland. *Eur J Clin Microbiol Infect Dis* 19: 790-31. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
24. Kremling A, Jahreis K, Lengeler JW, Gilles ED (2000) The organization of metabolic reaction networks: a signal-oriented approach to cellular models. *Metab Eng* 2: 1. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
25. Lenhard B, Wasserman WW (2002) TFBS: Computational framework for transcription factor binding site analysis. *Bioinformatics* 18: 1135-6. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
26. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, et al. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30: 325. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
27. Liu X, Brutlag DL, Liu JS (2001) Bioprosppector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pa Symp Biocomput* 127-38. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
28. Maeda S, Yoshida H, Ikenoue T, Ogura K, Kanai F, et al. (1999) Structure of cag pathogenicity island in Japanese *Helicobacter pylori* isolates. *Gut* 44: 307-8. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
29. Matysiak-Budnik T, Megraud F (1994) *Helicobacter pylori* in eastern European countries: what is the current status? *Gut* 35: 1683-6. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
30. McGurie AM, Hughes JD, Church GM (2001) conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Research* 10: 744-757. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
31. Merrell DS, Thompson LJ, Kim CC, Mitchell H, Tompkins LS, et al. (2003) Growth phase-dependent response of *Helicobacter pylori* to iron starvation. *Infect Immun* 71: 6510-25. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
32. Ohata H, Kitauchi S, Yoshimura N, Mugitani K, Iwane M, et al. (2004) Progression of chronic atrophic gastritis associated with *Helicobacter pylori* infection increases risk of gastric cancer. *Int J Cancer* 109: 138-43. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
33. Parkin DM (2001) Global cancer statistics in the year 2000. *Lancet Oncol* 2: 533-43. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
34. Pribnow D (1975) Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci* 72: 784-788. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
35. Price MN, Huang KH, Alm EJ, Arkin AP (2005) A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res* 33: 880-92. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
36. Salgado H, Santos-Zavaleta A, Gama-Castro S, Millán-Zárate D, Blattner FR, et al. (2000) RegulonDB (version 3.0): Transcriptional regulation and operon organization in *Escherichia coli*. *Nucleic Acids Res* 28: 65-67. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
37. Schena M, Shalon D, Heller R, Chai A, Brown PO, et al. (1996) Parallel human genome analysis: microarray-based gene expression monitoring of 1000 genes. *Proc Natl Acad Sci* 93: 10614-9. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
38. Spellman P, Sherlock G, Zhang M, Iyer V, Anders K, et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 9: 3273-3297. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
39. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, et al. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci* 96: 2907-12. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)

40. Thomas-Chollier M, Sand O, Turatsinze JV, Janky R, Defrance M, et al. (2008) RSAT: regulatory sequence analysis tools *Nucleic Acids Res* 36: W119-27. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
41. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, et al. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539-547. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
42. Ulm R, Baumann A, Oravec A, Mate Z, Adam E, et al. (2004) Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of *Arabidopsis*. *Proc Natl Acad Sci* 101: 1397-402. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
43. van-Helden J (2003) Regulatory sequence analysis tools. *Nucleic Acids Res* 31: 3593-6. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
44. van-Helden J, Rios AF, Collado-Vides J (2000) Discovering regulatory elements in intergenic sequences by analysis of spaced dyads. *Nucleic Acids Res* 28: 1808-18. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
45. van-Noort V, Snel B, Huynen MA (2003) Predicting gene function by conserved co-expression. *Trends Genet* 19: 238-242. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
46. Wen X, Furham S, Michaels GS, Carr DB, Smith S, et al. (1998) Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci* 95: 334-339. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
47. Yeung KY, Medvedovic M, Bumgarner RE (2004) From co-expression to co-regulation: how many microarray experiments do we need? *Genome Biol* 5: R48 Epub 28. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
48. Yoon S, Hur C, Kang H, Kim Y, Oh T, et al. (2005) A computational approach for identifying Pathogenicity islands in prokaryotic genomes. *BMC Bioinformatics* 6: 184. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)