Inducible gene expression systems are favored over stable expression systems in a wide variety of basic and applied research areas, including functional genomics, gene therapy, tissue engineering, biopharmaceutical protein production and drug discovery. This is because they are mostly reversible and thus more flexible to use. Furthermore, compared to constitutive expression, they generally exhibit a higher efficiency and have fewer side effects, such as cell death and delayed growth or development. Empowered by decades of development of inducible gene expression systems, researchers can now efficiently activate or suppress any gene, temporarily and quantitively at will, depending on experimental requirements and designs. Here, we review a number of most commonly used mammalian inducible expression systems and provide basic standards and criteria for the selection of the most suitable one.

We report the development of an advanced system for transfer and expression of exogenous genes in mammalian cells based on Moloney murine leukemia virus (Mo MuLV). Extensive deletion/mutagenesis analysis to identify cis-acting signals involved in virus transmission has led to the design of a family of novel, highly efficient retroviral vectors and a partner helper-free packaging cell line. The pBabe retroviral vector constructs transmit inserted genes at high titres and express them from the Mo MuLV Long Terminal Repeat (LTR). Each of these vectors has been constructed with one of four different dominantly acting selectable markers, allowing the growth of infected mammalian cells in the presence of G418, hygromycin B, bleomycin/phleomycin or puromycin, respectively. The high titre ecotropic helper free packaging cell line, omega E, was designed in conjunction with the pBabe vectors to reduce the risk of generation of wild type Mo MuLV via homologous recombination events. The omega E cell line was generated with separate gagpol and ecotropic env expression constructs with minimal sequence overlap and decreased sequence homology achieved by 'codon wobbling'. Homologous env coding sequences were deleted from the pBabe vectors without diminishing recombinant vector titre. Together, the pBabe vectors and omega E cell line should prove useful in experiments where highest frequencies of gene transfer, or concomitant expression of several different genes within a single cell are required with minimal risk of helper virus contamination.

The Drug-Gene Interaction Database (DGIdb, www.dgidb.org) is a web resource that consolidates disparate data sources describing drug-gene interactions and gene druggability. It provides an intuitive graphical user interface and a documented application programming interface (API) for querying these data. DGIdb was assembled through an extensive manual curation effort, reflecting the combined information of twenty-seven sources. For DGIdb 2.0, substantial updates have been made to increase content and improve its usefulness as a resource for mining clinically actionable drug targets. Specifically, nine new sources of drug-gene interactions have been added, including seven resources specifically focused on interactions linked to clinical trials. These additions have more than doubled the overall count of drug-gene interactions. The total number of druggable gene claims has also increased by 30%. Importantly, a majority of the unrestricted, publicly-accessible sources used in DGIdb are now automatically updated on a weekly basis, providing the most current information for these sources. Finally, a new web view and API have been developed to allow searching for interactions by drug identifiers to complement existing gene-based search functionality. With these updates, DGIdb represents a comprehensive and user friendly tool for mining the druggable genome for precision medicine hypothesis generation.

The drug-gene interaction database (DGIdb, www.dgidb.org) consolidates, organizes and presents drug-gene interactions and gene druggability information from papers, databases and web resources. DGIdb normalizes content from 30 disparate sources and allows for user-friendly advanced browsing, searching and filtering for ease of access through an intuitive web user interface, application programming interface (API) and public cloud-based server image. DGIdb v3.0 represents a major update of the database. Nine of the previously included 24 sources were updated. Six new resources were added, bringing the total number of sources to 30. These updates and additions of sources have cumulatively resulted in 56 309 interaction claims. This has also substantially expanded the comprehensive catalogue of druggable genes and anti-neoplastic drug-gene interactions included in the

DGIdb. Along with these content updates, v3.0 has received a major overhaul of its codebase, including an updated user interface, preset interaction search filters, consolidation of interaction information into interaction groups, greatly improved search response times and upgrading the underlying web application framework. In addition, the expanded API features new endpoints which allow users to extract more detailed information about queried drugs, genes and drug-gene interactions, including listings of PubMed IDs, interaction type and other interaction metadata.

Molecular imaging has played an important role in the noninvasive exploration of multiple biological processes. Reporter gene imaging is a key part of molecular imaging. By combining with a reporter probe, a reporter protein can induce the accumulation of specific signals that are detectable by an imaging device to provide indirect information of reporter gene expression in living subjects. There are many types of reporter genes and each corresponding imaging technique has its own advantages and drawbacks. Fused reporter genes or single reporter genes with products detectable by multiple imaging modalities can compensate for the disadvantages and potentiate the advantages of each modality. Reporter gene multimodality imaging could be applied to trace implanted cells, monitor gene therapy, assess endogenous molecular events, screen drugs, etc. Although several types of multimodality imaging apparatus and multimodality reporter genes are available, more sophisticated detectors and multimodality reporter gene systems are needed.

Genomic data interpretation often requires analyses that move from a gene-by-gene focus to a focus on sets of genes that are associated with biological phenomena such as molecular processes, phenotypes, diseases, drug interactions or environmental conditions. Unique challenges exist in the curation of gene sets beyond the challenges in curation of individual genes. Here we highlight a literature curation workflow whereby gene sets are curated from peer-reviewed published data into GeneWeaver (GW), a data repository and analysis platform. We describe the system features that allow for a flexible yet precise curation procedure. We illustrate the value of curation by gene sets through analysis of independently curated sets that relate to the integrated stress response, showing that sets curated from independent sources all share significant Jaccard similarity. A suite of reproducible analysis tools is provided in GW as services to carry out interactive functional investigation of user-submitted gene sets within the context of over 150 000 gene sets constructed from publicly available resources and published gene lists. A curation interface supports the ability of users to design and maintain curation workflows of gene sets, including assigning, reviewing and releasing gene sets within a curation project context.

Chemoresistance genes, initially considered to be a major impediment to the successful treatment of cancer, may become useful tools for gene therapy of cancer and of genetically determined disorders. Various target cells are rendered resistant to anticancer drugs by transfer of chemoresistance genes encoding P-glycoprotein, the multidrug resistance-associated protein-transporter, dihydrofolate reductase, glutathione-S-transferase, O6-alkylguanine DNA alkyltransferase, or aldehyde reductase. These genes can be used for selection in vivo because of the pharmacology and pharmacokinetics of their substrates. In contrast, several other selectable marker genes conferring resistance to substrates like neomycin or hygromycin can only be utilized in tissue culture. Possible applications for chemoresistance genes include protection of bone marrow and other organs from adverse effects caused by the toxicity of chemotherapy. Strategies have also been developed to introduce and overexpress nonselectable genes in target cells by cotransduction with chemoresistance genes. Thereby expression of both transgenes can be increased following selection with drugs. Moreover, treatment with chemotherapeutic agents should restore transgene expression when or if expression levels decrease after several weeks or months. This approach may improve the efficacy of somatic gene therapy of hematopoietic disorders which is hampered by low or unstable gene expression in progenitor cells. In this article we review preclinical studies in tissue culture and animal models, and ongoing clinical trials on transfer of chemoresistance genes to hematopoietic precursor cells of cancer patients.

The problems associated with gene identification and the prediction of gene structure in DNA sequences have been the focus of increased attention over the past few years with the recent acquisition by large-scale sequencing projects of an immense amount of genome data. A variety of prediction programs have been developed in order to address these problems. This paper presents a review of the computational approaches and gene-finders used commonly for gene prediction in eukaryotic genomes. Two approaches, in general, have been adopted for this purpose: similarity-based and ab initio techniques. The information gleaned from these methods is then combined via a variety of algorithms, including Dynamic Programming (DP) or the Hidden Markov Model (HMM), and then used for gene prediction from the genomic sequences.

Although evidence indicates that drug target genes share some common evolutionary features, there have been few studies analyzing evolutionary features of drug targets from an overall level. Therefore, we conducted an analysis which aimed to investigate the evolutionary characteristics of drug target genes. We compared the evolutionary conservation between human drug target genes and non-target genes by combining both the evolutionary features and network topological properties in human protein-protein interaction network. The evolution rate, conservation score and the percentage of orthologous genes of 21 species were included in our study. Meanwhile, four topological features including the average shortest path length, betweenness centrality, clustering coefficient and degree were considered for comparison analysis. Then we got four results as following: compared with non-drug target genes, 1) drug target genes had lower evolutionary rates; 2) drug target genes had higher conservation scores; 3) drug target genes had higher percentages of orthologous genes and 4) drug target genes had a tighter network structure including higher degrees, betweenness centrality, clustering coefficients and lower average shortest path lengths. These results demonstrate that drug target genes are more evolutionarily conserved than non-drug target genes. We hope that our study will provide valuable information for other researchers who are interested in evolutionary conservation of drug targets.

The panel of 60 human cancer cell lines (the NCI-60) assembled by the National Cancer Institute for anticancer drug discovery is a widely used resource. The NCI-60 has been characterized pharmacologically and at the molecular level more extensively than any other set of cell lines. However, no systematic mutation analysis of genes causally implicated in oncogenesis has been reported. This study reports the sequence analysis of 24 known cancer genes in the NCI-60 and an assessment of 4 of the 24 genes for homozygous deletions. One hundred thirty-seven oncogenic mutations were identified in 14 (APC, BRAF, CDKN2, CTNNB1, HRAS, KRAS, NRAS, SMAD4, PIK3CA, PTEN, RB1, STK11, TP53, and VHL) of the 24 genes. All lines have at least one mutation among the cancer genes examined, with most lines (73%) having more than one. Identification of those cancer genes mutated in the NCI-60, in combination with pharmacologic and molecular profiles of the cells, will allow for more informed interpretation of anticancer agent screening and will enhance the use of the NCI-60 cell lines for molecularly targeted screens.

Chemoprevention has been shown to be an extremely promising approach to the prevention of invasive cancer. Through the identification of chemopreventive agents that inhibit or reverse the process of carcinogenesis, new strategies of early intervention can be developed for patients at high risk that potentially prevent the onset of invasive and metastatic phases of cancer. This articles reviews the present efforts in chemoprevention research, including the identification of promising agents, screening, and preclinical and clinical evaluations.

The genetics of variable drug response appears to be a more tractable complex trait than common disease predisposition. This has implications for prioritizing research and for experimental design, and in particular argues for extensive use of candidate gene based approaches in pharmacogenetic association studies. Eventually, when whole genome scanning becomes feasible, it may be appropriate to consider weighting schemes that assign higher prior probabilities of variants in genes related to the mode of action of metabolism of medicines.

Gene therapy has the potential to provide cancer treatments based on novel mechanisms of action with potentially low toxicities. This therapy may provide more effective control of locoregional recurrence in diseases like non-small-cell lung cancer (NSCLC) as well as systemic control of micrometastases. Despite current limitations, retroviral and adenoviral vectors can, in certain circumstances, provide an effective means of delivering therapeutic genes to tumor cells. Although multiple genes are involved in carcinogenesis, mutations of the p53 gene are the most frequent abnormality identified in human tumors. Preclinical studies both in vitro and in vivo have shown that restoring p53 function can induce apoptosis in cancer cells. High levels of p53 expression and DNA-damaging agents like cisplatin (Platinol) and ionizing radiation work synergistically to induce apoptosis in cancer cells. Phase I clinical trials now show that p53 gene replacement therapy using both retroviral and adenoviral vectors is feasible and safe. In addition, p53 gene replacement therapy induces tumor regression in patients with advanced NSCLC and in those with recurrent head and neck cancer. This article describes various gene therapy strategies under investigation, reviews preclinical data that provide a rationale for the gene replacement approach, and discusses the clinical trial data available to date.

Candida albicans is the primary fungal pathogen of humans. Despite the need for novel drugs to combat fungal infections [Sobel, J.D. (2000) Clin Infectious Dis 30: 652], antifungal drug discovery is currently limited by both the availability of suitable drug targets and assays to screen corresponding targets. A functional genomics approach based on the diploid C. albicans genome sequence, termed GRACETM (gene replacement and conditional expression), was used to assess gene essentiality through a combination of gene replacement and conditional gene expression. In a systematic application of this approach, we identify 567 essential genes in C. albicans. Interestingly, evaluating the conditional phenotype of all identifiable C. albicans homologues of the Saccharomyces cerevisiae essential gene set [Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., et al. (2002) Nature 418: 387-391] by GRACE revealed only 61% to be essential in C. albicans, emphasizing the importance of performing such studies directly within the pathogen. Construction of this conditional mutant strain collection facilitates large-scale examination of terminal phenotypes of essential genes. This information enables preferred drug targets to be selected from the C. albicans essential gene set by phenotypic information derived both in vitro, such as cidal versus static terminal phenotypes, as well as in vivo through virulence studies using conditional strains in an animal model of infection. In addition, the combination of phenotypic and bioinformatic analyses further improves drug target selection from the C. albicans essential gene set, and their respective conditional mutant strains may be directly used as sensitive whole-cell assays for drug screening.

The protective effect of half of the tested botanical varieties via the activation of EpRE-mediated gene expression was confirmed. The study also provided an example of how in vitro bioassays can be combined with LC-MS and the automated chemical annotation software MAGMa, to identify biologically active constituents in complex botanical extracts.

This article introduces DoBISCUIT (Database of BIoSynthesis clusters CUrated and InTegrated, http://www.bio.nite.go.jp/pks/), a literature-based, manually curated database of gene clusters for secondary metabolite biosynthesis. Bacterial secondary metabolites often show pharmacologically important activities and can serve as lead compounds and/or candidates for drug development. Biosynthesis of each secondary metabolite is catalyzed by a number of enzymes, usually encoded by a gene cluster. Although many scientific papers describe such gene clusters, the gene information is not always described in a comprehensive manner and the related information is rarely integrated. DoBISCUIT integrates the latest literature information and provides standardized gene/module/domain descriptions related to the gene clusters.

The basic region-leucine zipper (bZIP) transcription factors (TFs) form homodimers and heterodimers via the coil-coil region. The bZIP dimerization network influences gene expression across plant development and in response to a range of environmental stresses. The recent release of the most comprehensive potato reference genome was used to identify 80 *StbZIP* genes and to characterize their

gene structure, phylogenetic relationships, and gene expression profiles. The *StbZIP* genes have undergone 22 segmental and one tandem duplication events. Ka/Ks analysis suggested that most duplications experienced purifying selection. Amino acid sequence alignments and phylogenetic comparisons made with the Arabidopsis *bZIP* family were used to assign the *StbZIP* genes to functional groups based on the Arabidopsis orthologs. The patterns of introns and exons were conserved within the assigned functional groups which are supportive of the phylogeny and evidence of a common progenitor. Inspection of the leucine repeat heptads within the bZIP domains identified a pattern of attractive pairs favoring homodimerization, and repulsive pairs favoring heterodimerization. These patterns of attractive and repulsive heptads were similar within each functional group for Arabidopsis and *S. tuberosum* orthologs. High-throughput RNA-seq data indicated the most highly expressed and repressed genes that might play significant roles in tissue growth and development, abiotic stress response, and response to pathogens including *Potato virus X*. These data provide useful information for further functional analysis of the *StbZIP* gene family and their potential applications in crop improvement.

With sequence analysis of the human genome well underway, there is an increasingly urgent challenge to understand the fundamental function and interplay of genes that build and maintain an organism. Several approaches will be critical for interpreting gene function, including random cDNA sequencing, expression profiling in different tissues, genetic analysis of human or model organism phenotypes, and creation of transgenic or "knockout" animals. Traditional gene-trapping approaches, in which genes are randomly disrupted with DNA elements inserted throughout the genome, have been used to generate large numbers of mutant organisms for genetic analysis. Recent modifications of gene-trapping methods and their increased use in mammalian systems are likely to result in a wealth of new information on gene function. Various trapping strategies allow genes to be segregated based on criteria like the specific subcellular location of an encoded protein, the tissue expression profile, or responsiveness to specific stimuli. Genome-wide gene-trapping strategies, which integrate gene discovery and expression profiling, can be applied in a massively parallel format to produce living assays for drug discovery.

Fungal diseases represent a major burden to health care globally. As with other pathogenic microbes, there is a limited number of agents suitable for use in treating fungal diseases, and resistance to these agents can develop rapidly. Cryptococcus neoformans is a basidiomycete fungus that causes cryptococcosis worldwide in both immunocompromised and healthy individuals. As a basidiomycete, it diverged from other common pathogenic or model ascomycete fungi more than 500 million years ago. Here, we report C. neoformans genes that are essential for viability as identified through forward and reverse genetic approaches, using an engineered diploid strain and genetic segregation after meiosis. The forward genetic approach generated random insertional mutants in the diploid strain, the induction of meiosis and sporulation, and selection for haploid cells with counterselection of the insertion event. More than 2,500 mutants were analyzed, and transfer DNA (T-DNA) insertions in several genes required for viability were identified. The genes include those encoding the thioredoxin reductase (Trr1), a ribosome assembly factor (Rsa4), an mRNA-capping component (Cet1), and others. For targeted gene replacement, the C. neoformans homologs of 35 genes required for viability in ascomycete fungi were disrupted, meiosis and sporulation were induced, and haploid progeny were evaluated for their ability to grow on selective media. Twenty-one (60%) were found to be required for viability in C. neoformans. These genes are involved in mitochondrial translation, ergosterol biosynthesis, and RNA-related functions. The heterozygous diploid mutants were evaluated for haploinsufficiency on a number of perturbing agents and drugs, revealing phenotypes due to the loss of one copy of an essential gene in C. neoformans. This study expands the knowledge of the essential genes in fungi using a basidiomycete as a model organism. Genes that have no mammalian homologs and are essential in both Cryptococcus and ascomycete human pathogens would be ideal for the development of antifungal drugs with broad-spectrum activity.

An increased number of copies of specific genes may offer an advantage to cells when they grow in restrictive conditions such as in the presence of toxic drugs, or in a tumor. Three mathematical models

of gene amplification and deamplification are proposed to describe the kinetics of unstable phenotypes of cells with amplified genes. The models differ in details but all assume probabilistic mechanisms of increase and decrease in gene copy number per cell (gene amplification/deamplification). Analysis of the models indicates that a stable distribution of numbers of copies of genes per cell, observed experimentally, exists only if the probability of deamplification exceeds the probability of amplification. The models are fitted to published data on the loss of methotrexate resistance in cultured cell lines, due to the loss of amplified dihydrofolate reductase gene. For two mouse cell lines unstably resistant to methotrexate the probabilities of amplification and deamplification of the dihydrofolate reductase gene on double minute chromosomes are estimated to be approximately 2% and 10%, respectively. These probabilities are much higher than widely presumed. The models explain the gradual disappearance of the resistant phenotype when selective pressure is withdrawn, by postulating that the rate of deamplification exceeds the rate of amplification. Thus it is not necessary to invoke a growth advantage of nonresistant cells which has been the standard explanation. For another analogous process, the loss of double minute chromosomes containing the myc oncogene from SEWA tumor cells, the growth advantage model does seem to be superior to the amplification and deamplification model. In a more theoretical section of the paper, it is demonstrated that gene amplification/deamplification can result in reduction to homozygosity, such as is observed in some tumors. Other applications are discussed.

In a screen of 1,000 consecutively ascertained families, we recently found that mutations in the gene *RPGR* are the third most common cause of all inherited retinal disease. As the two most frequent disease-causing genes, *ABCA4* and *USH2A*, are far too large to fit into clinically relevant adeno-associated virus (AAV) vectors, *RPGR* is an obvious early target for AAV-based ocular gene therapy. In generating plasmids for this application, we discovered that those containing wild-type *RPGR* sequence, which includes the highly repetitive low complexity region ORF15, were extremely unstable (*i.e.*, they showed consistent accumulation of genomic changes during plasmid propagation). To develop a stable *RPGR* gene transfer vector, we used a bioinformatics approach to identify predicted regions of genomic instability within ORF15 (*i.e.*, potential non-B DNA conformations). Synonymous substitutions were made in these regions to reduce the repetitiveness and increase the molecular stability while leaving the encoded amino acid sequence unchanged. The resulting construct was subsequently packaged into AAV serotype 5, and the ability to drive transcript expression and functional protein production was demonstrated via subretinal injection in rat and pull-down assays, respectively. By making synonymous substitutions within the repetitive region of *RPGR*, we were able to stabilize the plasmid and subsequently generate a clinical-grade gene transfer vector (IA-RPGR). Following subretinal injection in rat, we demonstrated that the augmented transcript was expressed at levels similar to wild-type constructs. By performing *in vitro* pull-down experiments, we were able to show that IA-RPGR protein product retained normal protein binding properties (*i.e.*, analysis revealed normal binding to PDE6D, INPP5E, and RPGRIP1L). In summary, we have generated a stable *RPGR* gene transfer vector capable of producing functional RPGR protein, which will facilitate safety and toxicity studies required for progression to an Investigational New Drug application.

The Saccharomyces cerevisiae gene PDR1, responsible for pleiotropic drug resistance, was isolated from a genomic DNA cosmid library by hybridization to the flanking LEU1 gene, followed by subcloning the drug-sensitive phenotype into the transformed pdr1-1, pdr1-2, and pdr1-3 drug-resistant mutants. A RNA molecule of 3.5 kilobases was identified as the PDR1 transcript. The nucleotide sequence of the complementing DNA fragment contained a 3192-nucleotide open reading frame. Disruption of the pdr1 and PDR1 genes restored or increased drug sensitivity. Analysis of the PDR1 deduced amino acid sequence revealed several homologies to four different regulatory proteins involved in the control of gene expression, including a cysteine-rich motif suggested to be a metal-binding domain for DNA recognition. A model is proposed of a general transcriptional control by PDR1 of several target genes encoding proteins from plasma, mitochondria, and possibly other permeability barriers.

Precise control of gene expression is a powerful method to elucidate biological function, and protein overexpression is an important tool for industry and biochemistry. Expression of the Neurospora crassa tcu-1 gene (NCU00830), encoding a high-affinity copper transporter, is tightly controlled by copper availability. Excess copper represses, and copper depletion, via the use of a copper chelator, activates expression. The kinetics of induction and repression of tcu-1 are rapid, and the effects are long lived. We constructed a plasmid carrying the bar gene (for glufosinate selection) fused to the tcu-1 promoter. This plasmid permits the generation of DNA fragments that can direct integration of Ptcu-1 into any desired locus. We use this strategy to integrate Ptcu-1 in front of wc-1, a circadian oscillator and photoreceptor gene. The addition of excess copper to the Ptcu-1::wc-1 strain phenocopies a Δwc-1 strain, and the addition of the copper chelator, bathocuproinedisulfonic acid, phenocopies a wc-1 overexpression strain. To test whether copper repression can recapitulate the loss of viability that an essential gene knockout causes, we placed Ptcu-1 upstream of the essential gene, hpt-1. The addition of excess copper drastically reduced the growth rate as expected. Thus, this strategy will be useful to probe the biological function of any N. crassa gene through controlled expression.

Cannabinoid receptor-1 (CB1) represents a potential drug target against conditions that include obesity and substance abuse. However, drug trials targeting CB1 (encoded by the CNR1 gene) have been compromised by differences in patient response. Toward addressing the hypothesis that genetic changes within the regulatory regions controlling CNR1 expression contribute to these differences, we characterized the effects of disease-associated allelic variation within a conserved regulatory sequence (ECR1) in CNR1 intron 2 that had previously been shown to modulate cannabinoid response, alcohol intake, and anxiety-like behavior. We used primary cell analysis of reporters carrying different allelic variants of the human ECR1 and found that human-specific C-allele variants of ECR1 (ECR1(C)) drove higher levels of CNR1prom activity in primary hippocampal cells than did the ancestral T-allele and demonstrated a differential response to CB1 agonism. We further demonstrate a role for the AP-1 transcription factor in driving higher ECR1(C) activity and evidence that the ancestral t-allele variant of ECR1 interacted with higher affinity with the insulator binding factor CTCF. The cell-specific approaches used in our study represent an important step in gaining a mechanistic understanding of the roles of noncoding polymorphic variation in disease and in the increasingly important field of cannabinoid pharmacogenetics.

Identification of differentially expressed genes from transcriptomic studies is one of the most common mechanisms to identify tumor biomarkers. This approach however is not well suited to identify interaction between genes whose protein products potentially influence each other, which limits its power to identify molecular wiring of tumour cells dictating response to a drug. Due to the fact that signal transduction pathways are not linear and highly interlinked, the biological response they drive may be better described by the relative amount of their components and their functional relationships than by their individual, absolute expression.

Over the past decade, bacterial genome sequences have revealed an immense reservoir of biosynthetic gene clusters, sets of contiguous genes that have the potential to produce drugs or drug-like molecules. However, the majority of these gene clusters appear to be inactive for unknown reasons prompting terms such as "cryptic" or "silent" to describe them. Because natural products have been a major source of therapeutic molecules, methods that rationally activate these silent clusters would have a profound impact on drug discovery. Herein, a new strategy is outlined for awakening silent gene clusters using small molecule elicitors. In this method, a genetic reporter construct affords a facile read-out for activation of the silent cluster of interest, while high-throughput screening of small molecule libraries provides potential inducers. This approach was applied to two cryptic gene clusters in the pathogenic model Burkholderia thailandensis. The results not only demonstrate a prominent activation of these two clusters, but also reveal that the majority of elicitors are themselves antibiotics, most in common clinical use. Antibiotics, which kill B. thailandensis at high concentrations, act as inducers of secondary metabolism at low concentrations. One of these antibiotics, trimethoprim, served as a global activator of secondary metabolism by inducing at least five biosynthetic pathways. Further application of this

strategy promises to uncover the regulatory networks that activate silent gene clusters while at the same time providing access to the vast array of cryptic molecules found in bacteria.

We propose the hypothesis that for a particular type of cancer there exists a key pair of oncogene (OCG) and tumor suppressor gene (TSG) that is normally involved in strong stabilizing negative feedback loops (nFBLs) of molecular interactions, and it is these interactions that are sufficiently perturbed during cancer development. These nFBLs are thought to regulate oncogenic positive feedback loops (pFBLs) that are often required for the normal cellular functions of oncogenes. Examples given in this paper are the pairs of MYC and p53, KRAS and INK4A, and E2F1 and miR-17-92. We propose dynamical models of the aforementioned OCG-TSG interactions and derive stability conditions of the steady states in terms of strengths of cycles in the qualitative interaction network. Although these conditions are restricted to predictions of local stability, their simple linear expressions in terms of competing nFBLs and pFBLs make them intuitive and practical guides for experimentalists aiming to discover drug targets and stabilize cancer networks.

This study demonstrates the feasibility of combining two discrete drug-regulated expression systems in a temporally sequential cascade, without loss of dynamic range of signal induction. The efficient layering of control levels allowed by this combination of elements provides the potential for the generation of complex control circuitry that may advance ability to regulate gene expression in vivo.

Although the emergence of bacterial drug resistance is of great concern to the scientific community, few studies have evaluated this phenomenon systematically in fungi by using genome-wide datasets. In the present study, we assembled a large compendium of Saccharomyces cerevisiae chemical genetic data to study the evolution of multidrug resistance genes (MDRs) in the fungal lineage. We found that MDRs typically emerge in widely conserved families, most of which containing homologs from pathogenic fungi, such as Candida albicans and Coccidioides immitis, which could favor the evolution of drug resistance in those species. By integrating data from chemical genetics with protein family conservation, genetic and protein interactions, we found that gene families rarely have more than one MDR, indicating that paralogs evolve asymmetrically with regard to multidrug resistance roles. Furthermore, MDRs have more genetic and protein interaction partners than non-MDRs, supporting their participation in complex biochemical systems underlying the tolerance to multiple bioactive molecules. MDRs share more chemical genetic interactions with other MDRs than with non-MDRs, regardless of their evolutionary affinity. These results suggest the existence of an intricate system involved in the global drug tolerance phenotypes. Finally, MDRs are more likely to be hit repeatedly by mutations in laboratory evolution experiments, indicating that they have great adaptive potential. The results presented here not only reveal the main genomic features underlying the evolution of MDRs, but also shed light on the gene families from which drug resistance is more likely to emerge in fungi.

Tumors often have DNA repair defects, suggesting additional inhibition of other DNA repair pathways in tumors may lead to synthetic lethality. Accumulating data demonstrate that DNA repair-defective tumors, in particular homologous recombination (HR), are highly sensitive to DNA-damaging agents. Thus, HR-defective tumors exhibit potential vulnerability to the synthetic lethality approach, which may lead to new therapeutic strategies. It is well known that poly (adenosine diphosphate (ADP)-ribose) polymerase (PARP) inhibitors show the synthetically lethal effect in tumors defective in BRCA1 or BRCA2 genes encoded proteins that are required for efficient HR. In this review, we summarize the strategies of targeting DNA repair pathways and other DNA metabolic functions to cause synthetic lethality in HR-defective tumor cells.

Compiling a comprehensive list of cancer driver genes is imperative for oncology diagnostics and drug development. While driver genes are typically discovered by analysis of tumor genomes, infrequently mutated driver genes often evade detection due to limited sample sizes. Here, we address sample size limitations by integrating tumor genomics data with a wide spectrum of gene-specific properties to search for rare drivers, functionally classify them, and detect features characteristic of driver genes. We

show that our approach, CAnceR geNe similarity-based Annotator and Finder (CARNAF), enables detection of potentially novel drivers that eluded over a dozen pan-cancer/multi-tumor type studies. In particular, feature analysis reveals a highly concentrated pool of known and putative tumor suppressors among the <1% of genes that encode very large, chromatin-regulating proteins. Thus, our study highlights the need for deeper characterization of very large, epigenetic regulators in the context of cancer causality.

Discovery of natural products that possess novel chemical structures and pharmaceutical activities increases opportunities of drug development. Filamentous fungi have been recognized as an attractive source for pharmaceutically beneficial natural products. Genome sequencing innovation represented by Next-generation sequencer opened fungal genomes one after another, suggesting that one fungal strain has far more biosynthetic gene clusters than that are estimated from the number of previously isolated natural products. In addition, bioinformatics analyses have indicated that most biosynthetic gene clusters are silent under laboratory culture conditions and there are a huge number of natural products hidden in the fungal genome. Therefore, we focused on those silent biosynthetic gene clusters as a potential source for novel natural products and developed methods to activate silent biosynthetic gene clusters by using low molecular weight molecules. In this review, we describe on discovery of novel natural products through activating fungal silent biosynthesis by addition of epigenetic modifiers and plant hormones.

In this review we describe the expanding repertoire of molecular tools with which to study gene function in Leishmania. Specifically we review the tools available for studying functions of essential genes, such as plasmid shuffle and DiCre, as well as the rapidly expanding portfolio of available CRISPR/Cas9 approaches for large scale gene knockout and endogenous tagging. We include detail on approaches that allow the direct manipulation of RNA using RNAi and protein levels via Tet or DiCre induced overexpression and destabilization domain mediated degradation. The utilisation of current methods and the development of more advanced molecular tools will lead to greater understanding of the role of essential genes in the parasite and thereby more robust drug target validation, thereby paving the way for the development of novel therapeutics to treat this important disease.

There have been tremendous advances in our understanding of cancer from the application of molecular biology over the past decade. The disease is caused by a series of defects in the genes that accelerate growth--oncogenes--and those that slow down cellular turnover--tumour suppressor genes. The proteins they encode provide a promising hunting ground in which to design and test new anticancer drugs. Several treatment strategies are now under clinical trial entailing direct gene transfer. These include the use of gene marking to detect minimal residual disease, the production of novel cancer vaccines by the insertion of genes which uncloak cancer cells so making them visible to the host's immune system, the isolation and coupling of cancer specific molecular switches upstream of drug activating genes, and the correction of aberrant oncogenes or tumour suppressor genes. The issues in these approaches are likely to have a profound impact on the management of cancer patients as we enter the next century.

Chemically inducible gene switches that regulate expression of endogenous genes have multiple applications for basic gene expression research and gene therapy. Single-chain zinc-finger transcription factors that utilize either estrogen receptor homodimers or retinoid X receptor-alpha/ecdysone receptor heterodimers are shown here to be effective regulators of ICAM-1 and ErbB-2 transcription. Using activator (VP64) and repressor (Krüppel-associated box) domains to impart regulatory directionality, ICAM-1 was activated by 4.8-fold and repressed by 81% with the estrogen receptor-inducible transcription factors. ErbB-2 was activated by up to threefold and repressed by 84% with the retinoid X receptor-alpha/ecdysone receptor-inducible transcription factors. The dynamic range of these proteins was similar to the constitutive system and showed negligible basal regulation when ligand was not present. We have also demonstrated that the regulation imposed by these inducible transcription factors is dose dependent, sustainable for at least 11 days and reversible upon cessation of drug treatment. Importantly, these proteins can be used in conjunction with each other with no detectable

overlap of activity enabling concurrent and temporal regulation of multiple genes within the same cell. Thus, these chemically inducible transcription factors are valuable tools for spatiotemporal control of gene expression that should prove valuable for research and gene therapy applications.

Two genes are called synthetic lethal (SL) if mutation of either alone is not lethal, but mutation of both leads to death or a significant decrease in organism's fitness. The detection of SL gene pairs constitutes a promising alternative for anti-cancer therapy. As cancer cells exhibit a large number of mutations, the identification of these mutated genes' SL partners may provide specific anti-cancer drug candidates, with minor perturbations to the healthy cells. Since existent SL data is mainly restricted to yeast screenings, the road towards human SL candidates is limited to inference methods.

OGEE is an Online GEne Essentiality database. To enhance our understanding of the essentiality of genes, in OGEE we collected experimentally tested essential and non-essential genes, as well as associated gene properties known to contribute to gene essentiality. We focus on large-scale experiments, and complement our data with text-mining results. We organized tested genes into data sets according to their sources, and tagged those with variable essentiality statuses across data sets as conditionally essential genes, intending to highlight the complex interplay between gene functions and environments/experimental perturbations. Developments since the last public release include increased numbers of species and gene essentiality data sets, inclusion of non-coding essential sequences and genes with intermediate essentiality statuses. In addition, we included 16 essentiality data sets from cancer cell lines, corresponding to 9 human cancers; with OGEE, users can easily explore the shared and differentially essential genes within and between cancer types. These genes, especially those derived from cell lines that are similar to tumor samples, could reveal the oncogenic drivers, paralogous gene expression pattern and chromosomal structure of the corresponding cancer types, and can be further screened to identify targets for cancer therapy and/or new drug development

Despite the examples of protein evolution via mutations in coding sequences, we have very limited understanding on gene network evolution via changes in cis-regulatory elements. Using the galactose network as a model, here we show how the regulatory promoters of the network contribute to the evolved network activity between two yeast species. In Saccharomyces cerevisiae, we combinatorially replace all regulatory network promoters by their counterparts from Saccharomyces paradoxus, measure the resulting network inducibility profiles, and model the results. Lowering relative strength of GAL80-mediated negative feedback by replacing GAL80 promoter is necessary and sufficient to have high network inducibility levels as in S. paradoxus. This is achieved by increasing OFF-to-ON phenotypic switching rates. Competitions performed among strains with or without the GAL80 promoter replacement show strong relationships between network inducibility and fitness. Our results support the hypothesis that gene network activity can evolve by optimizing the strength of negative-feedback regulation.

In a recent issue of Cell, Hughes and coworkers (Mnaimneh et al., 2004) provide a great leap forward in the analysis of essential yeast genes by constructing a strain set that expresses each essential gene from a tetracycline-regulatable promoter.

To quantify the transcriptional activity of NF-κB and to screen drugs related to the regulation of NF-κB activation, we constructed a recombinant plasmid through deleting the original CMV promoter of retrovirus vector pQCXIP and inserting the NF-κB enhancer and NanoLuc luciferase sequence into the vector. Then, using the recombinant plasmid we constructed a cell line in which the expression of NanoLuc luciferase (NLuc) was regulated by NF-κB. The inserted sequences were verified by restriction endonuclease digestion and sequencing. Tumor necrosis factor-α (TNF-α), an NF-κB activator, acted on the constructed NLuc cell line and leaded to the specific luciferase reaction. The luciferase reaction showed a fine time and dose dependence to the TNF-α stimulation, indicating the successful construction of the NF-κB regulated NLuc-expressing cell line. Besides, the NF-κB inhibitor, triptolide, reduced the expression of NLuc in a dose-dependent way. The constructed reporter system

in this study could be applied in the quantification of the NF-κB transcriptional activity and in the NF-κB regulation-related drug screening.

MicroRNAs (miRNA) are approximately 18 to 25 nucleotides in length and affect gene expression by silencing the translation of messenger RNAs. Because each miRNA regulates the expression of hundreds of different genes, miRNAs can function as master coordinators, efficiently regulating and coordinating multiple cellular pathways and processes. By coordinating the expression of multiple genes, miRNAs are responsible for fine-tuning the cell's most important processes, like the ones involved in cellular growth and proliferation. Dysregulation of miRNAs appears to play a fundamental role in the onset, progression and dissemination of many cancers, and replacement of downregulated miRNAs in tumor cells results in a positive therapeutic response. Thus, in theory, inhibition of a particular miRNA linked to cancer onset or progression can remove the inhibition of the translation of a therapeutic protein-and conversely, administration of a miRNA mimetic can boost the endogenous miRNA population repressing the translation of an oncogenic protein. Although several basic questions about their biologic principles still remain to be answered, and despite the fact that all data with respect to miRNAs and therapy are still at the preclinical level, many specific characteristics of miRNAs in combination with compelling therapeutic efficacy data have triggered the research community to start exploring the possibilities of using miRNAs as potential therapeutic candidates.

Although Aspergillus fumigatus is an important human fungal pathogen there are few expression systems available to study the contribution of specific genes to the growth and virulence of this opportunistic mould. Regulatable promoter systems based upon prokaryotic regulatory elements in the E. coli tetracycline-resistance operon have been successfully used to manipulate gene expression in several organisms, including mice, flies, plants, and yeast. However, the system has not yet been adapted for Aspergillus spp.

Mammalian carboxylesterase (CES or Ces) genes encode enzymes that participate in xenobiotic, drug, and lipid metabolism in the body and are members of at least five gene families. Tandem duplications have added more genes for some families, particularly for mouse and rat genomes, which has caused confusion in naming rodent Ces genes. This article describes a new nomenclature system for human, mouse, and rat carboxylesterase genes that identifies homolog gene families and allocates a unique name for each gene. The guidelines of human, mouse, and rat gene nomenclature committees were followed and "CES" (human) and "Ces" (mouse and rat) root symbols were used followed by the family number (e.g., human CES1). Where multiple genes were identified for a family or where a clash occurred with an existing gene name, a letter was added (e.g., human CES4A; mouse and rat Ces1a) that reflected gene relatedness among rodent species (e.g., mouse and rat Ces1a). Pseudogenes were named by adding "P" and a number to the human gene name (e.g., human CES1P1) or by using a new letter followed by ps for mouse and rat Ces pseudogenes (e.g., Ces2d-ps). Gene transcript isoforms were named by adding the GenBank accession ID to the gene symbol (e.g., human CES1_AB119995 or mouse Ces1e_BC019208). This nomenclature improves our understanding of human, mouse, and rat CES/Ces gene families and facilitates research into the structure, function, and evolution of these gene families. It also serves as a model for naming CES genes from other mammalian species.

Geminiviruses possess single-stranded, circular DNA genomes and control the transcription of their late genes, including BV1 of many bipartite begomoviruses, through transcriptional activation by the early expressing AC2 protein. DNA binding by AC2 is not sequence-specific; hence, the specificity of AC2 activation is thought to be conferred by plant transcription factors (TFs) recruited by AC2 in infected cells. However, the exact TFs AC2 recruits are not known for most viruses. Here, we report a systematic examination of the BV1 promoter (P_{BV1}) of the mungbean yellow mosaic virus (MYMV) for conserved promoter motifs. We found that MYMV P_{BV1} contains three abscisic acid (ABA)-responsive elements (ABREs) within its first 70 nucleotides. Deleting these ABREs, or mutating them all via site-directed mutagenesis, abolished the capacity of P_{BV1} to respond to AC2-mediated transcriptional activation. Furthermore, ABRE and other related ABA-responsive elements were

prevalent in more than a dozen Old World begomoviruses we inspected. Together, these findings suggest that ABA-responsive TFs may be recruited by AC2 to BV1 promoters of these viruses to confer specificity to AC2 activation. These observations are expected to guide the search for the actual TF(s), furthering our understanding of the mechanisms of AC2 action.

Identification of synthetic lethal interactions in cancer cells could offer promising new therapeutic targets. Large-scale functional genomic screening presents an opportunity to test large numbers of cancer synthetic lethal hypotheses. Methods enriching for candidate synthetic lethal targets in molecularly defined cancer cell lines can steer effective design of screening efforts. Loss of one partner of a synthetic lethal gene pair creates a dependency on the other, thus synthetic lethal gene pairs should never show simultaneous loss-of-function. We have developed a computational approach to mine large multi-omic cancer data sets and identify gene pairs with mutually exclusive loss-of-function. Since loss-of-function may not always be genetic, we look for deleterious mutations, gene deletion and/or loss of mRNA expression by bimodality defined with a novel algorithm BiSEp.

Gene networks in nanoscale are of nonlinear stochastic process. Time delays are common and substantial in these biochemical processes due to gene transcription, translation, posttranslation protein modification and diffusion. Molecular noises in gene networks come from intrinsic fluctuations, transmitted noise from upstream genes, and the global noise affecting all genes. Knowledge of molecular noise filtering and biochemical process delay compensation in gene networks is crucial to understand the signal processing in gene networks and the design of noise-tolerant and delay-robust gene circuits for synthetic biology.

DNA methylation is an epigenetic modification that mainly repress expression of genes essential during embryogenesis and development. There are key ATPase-dependent enzymes that read or write DNA methylation to remodel chromatin and regulate gene expression. Structural maintenance of chromosome hinge domain containing 1 (SMCHD1) is an architectural protein that regulates expression of numerous genes, some of which are imprinted, that are sensitive to DNA methylation. In addition, SMCHD1 germline mutations lead to developmental diseases; facioscapulohumoral muscular dystrophy (FSHD), bosma arhinia and micropthalmia (BAMS). Current evidence suggests that SMCHD1 functions through maintenance or de novo DNA methylation required for chromatin compaction. However, it is unclear if DNA methylation is also essential for genomic recruitment of SMCHD1 and its role as an architectural protein. We previously isolated SMCHD1 using a methylated DNA region from mouse pituitary growth hormone (Gh1) promoter, suggesting that methylation is required for SMCHD1 DNA binding. The goal of this study was to further understand DNA methylation directed role of SMCHD1 in regulating gene expression. Therefore, we profiled SMCHD1 genome wide occupancy in human neuroblastoma SH-SY5Y cells and evaluated if DNA methylation is required for SMCHD1 genomic binding by treating cells with the DNA demethylating reagent, 5-azacytidine (5-azaC).

Cancer is a complex disease where cancer cells express epigenetic and transcriptomic mechanisms to promote tumor initiation, progression, and survival. To extract relevant features from the 2019 Cancer Cell Line Encyclopedia (CCLE), a multi-layer nonnegative matrix factorization approach is used. We used relevant feature genes and DNA promoter regions to construct genomic interaction network to study gene-gene and gene-DNA promoter methylation relationships. Here, we identified a set of gene transcripts and methylated DNA promoter regions for different clusters, including one homogeneous lymphoid neoplasms cluster. In this cluster, we found different methylated transcription factors that affect transcriptional activation of EGFR and downstream interactions. Furthermore, the hippo-signaling pathway might not function properly because of DNA hypermethylation and low gene expression of both LATS2 and YAP1. Finally, we could identify a potential dysregulation of the CD28-CD86-CTLA4 axis. Characterizing the interaction of the epigenome and the transcriptome is vital for our understanding of cancer cell line behavior, not only for deepening insights into cancer-related processes but also for future disease treatment and drug development. Here we have identified potential

candidates that characterize cancer cell lines, which give insight into the development and progression of cancers.

Acclimation to high salt concentrations involves concerted changes in gene expression. For the majority of salt-regulated genes, the mechanism underlying the induction process is not known. The gene ggpS (sll1566), which encodes the glucosylglycerol-phosphate synthase responsible for the synthesis of the compatible solute glucosylglycerol (GG), is specifically induced by salt in the cyanobacterial model strain Synechocystis sp. strain PCC 6803. To identify mechanisms mediating this salt-specific gene regulation, the ggpS promoter was analyzed in more detail. 5' rapid amplification of cDNA ends (5'-RACE) experiments revealed that the adjacent open reading frame (ORF), which is annotated as unknown protein Ssl3076, overlaps with the transcriptional start site of the ggpS gene. Reporter gene expression analyses indicated an essential role for the intact ssl3076 gene in the salt-regulated transcription of a gfp reporter gene. Promoter fragments containing a mutated ssl3076 lost the salt regulation; similarly, a frameshift mutation in ssl3076 resulted in a high level of ggpS expression under low-salt conditions, thereby establishing this small ORF, named ggpR, as a negative regulator of ggpS. Interestingly, small ORFs were also found adjacent to ggpS genes in the genomes of other GG-accumulating cyanobacteria. These results suggest that the GgpR protein represses ggpS expression under low-salt conditions, whereas in salt-shocked and salt-acclimated cells a stress-proportional ggpS expression occurs, leading to GG accumulation.

Overexpression of MDR1, which encodes a membrane transport protein of the major facilitator superfamily, is one mechanism by which the human fungal pathogen Candida albicans can develop increased resistance to the antifungal drug fluconazole and other toxic compounds. In clinical C. albicans isolates, constitutive MDR1 overexpression is accompanied by the upregulation of other genes, but it is not known if these additional alterations are required for Mdr1p function and drug resistance. To investigate whether MDR1 overexpression is sufficient to confer a drug-resistant phenotype in C. albicans, we expressed the MDR1 gene from the strong ADH1 promoter in C. albicans laboratory strains that did not express the endogenous MDR1 gene as well as in a fluconazole-resistant clinical C. albicans isolate in which the endogenous MDR1 alleles had been deleted and in a matched fluconazole-susceptible isolate from the same patient. Forced MDR1 overexpression resulted in increased resistance to the putative Mdr1p substrates cerulenin and brefeldin A, and this resistance did not depend on the additional alterations which occurred during drug resistance development in the clinical isolates. In contrast, artificial expression of the MDR1 gene from the ADH1 promoter did not enhance or only slightly enhanced fluconazole resistance, presumably because Mdr1p expression levels in the transformants were considerably lower than those observed in the fluconazole-resistant clinical isolate. These results demonstrate that MDR1 overexpression in C. albicans is sufficient to confer resistance to some toxic compounds that are substrates of this efflux pump but that the degree of resistance depends on the Mdr1p expression level.

We report on the cloning and sequencing of the vanA gene cluster present in the glycopeptide-resistant clinical isolate Bacillus circulans VR0709 (R. Fontana, M. Ligozzi, C. Pedrotti, E. M. Padovani, and G. Cornaglia, Eur. J. Clin. Microbiol. Infect. Dis. 16:473-474, 1997). The presence of a vanA-related gene in VR0709 was demonstrated in a PCR assay which permitted the specific amplification of an internal segment of vanA. Southern blotting suggested that the vanA gene was located in the chromosome in a 7. 6-kb EcoRI fragment. DNA sequence analysis revealed the presence of all seven genes of the vanA cluster (vanR, vanS, vanH, vanA, vanX, vanY, and vanZ). The degree of identity between homologous proteins encoded by Tn1546 and the chromosome of B. circulans VR0709 ranged from 87 to 95%. Neither PCR nor Southern blotting with specific primers and probes, respectively, showed the presence of open reading frames (ORFs) 1 and 2 which encode the transposase and the resolvase of Tn1546, respectively, the transposon found to carry the vanA gene cluster in enterococci. Determination of the sequences of the flanking regions of the van gene cluster of B. circulans revealed perfect inverted repeats of 10 bp which delineated a 9.2-kb region containing the van gene cluster and an ORF which encoded a putative protein (178 residues) which displayed a low level of identity (28%)

to the resolvase of Tn1546. These results suggest that glycopeptide resistance in B. circulans VR0709 is associated with the acquisition of a vanA gene cluster which shows a high degree of homology with that of enterococci. In B. circulans, however, the cluster is not carried by Tn1546 and is borne by the chromosome.

New drug targets are urgently needed for parasites of socio-economic importance. Genes that are essential for parasite survival are highly desirable targets, but information on these genes is lacking, as gene knockouts or knockdowns are difficult to perform in many species of parasites. We examined the applicability of large-scale essentiality information from four model eukaryotes, Caenorhabditis elegans, Drosophila melanogaster, Mus musculus and Saccharomyces cerevisiae, to discover essential genes in each of their genomes. Parasite genes that lack orthologues in their host are desirable as selective targets, so we also examined prediction of essential genes within this subset.

The gene for an NAD-specific glutamate dehydrogenase (NAD-GDH) that is allosterically activated by NADP+ (non-substrate) was cloned, and its physical structure and nucleotide sequence was determined. The gene consists of 9 introns and 10 exons; the 10th and largest exon, which is 1863 nucleotides long, is at the 3'-end of the gene. The shortest exon of 33 base pairs is the first and is located at the 5'-end of the gene. The large exon is in perfect register along the complementary strand with a heat shock 70 (HSP)-like protein gene. The NAD-GDH gene is inducible with L-glutamine, just as the HSP 70-like protein gene (LéJohn, H.B., Cameron, L.E., Yang, B., MacBeath, G., Barker, D.S., and Williams, S.A. (1994) J. Biol. Chem. 269, 4513-4522). The phenomenon of anti-parallel coupling of two genes is named antisense gene pair. By Northern and Western blotting techniques, we obtained indirect evidence that the gene is expressed in vivo. The gene encodes a protein of M(r) 118,740 which consists of 1063 amino acid residues. The 5' and 3' borders of the gene display typical but unproven promoter motifs of CCAAT, TATAAT, and AAATAAAA polyadenylation signal bounded by a pyrimidine-rich transcription termination-type format. Restriction endonuclease site mapping of all the genomic clones isolated that carry most or all of the gene, and of the genome itself, gave hybridization patterns that are consistent with the interpretation that the organism, Achlya klebsiana, has only one form of the gene. 3'-End-labeling of a 5.2-kb XbaI DNA fragment (carrying the antisense gene pair) that was then asymmetrically cleaved to produce two single 3'-end-labeled pieces that were used as probes on L-glutamine-induced cell poly(A)+ RNA, showed that the end-labeled DNA equivalent to the HSP 70-like protein mRNA hybridized to a 3.4-kb transcript and the end-labeled DNA equivalent to the NAD-GDH mRNA hybridized to a 2.4-kb transcript.

Embryonal carcinoma (EC) cells offer an interesting model system for evaluating differentiation because the cells are pluripotent, thus resembling germ cells and embryonic stem cells, and because a number of agents have been defined that are capable of promoting the differentiation of these cells. This chapter examines how EC cells might be triggered to differentiate, with emphasis on retinoic acid because this compound is a potent, naturally occurring inducer that has been studied extensively in this system. The nature of alterations in gene expression during EC cell differentiation is reviewed from the perspective of evaluating whether these changes are likely to be responsible for, or a result of, the differentiation event. Finally, we consider in molecular terms why EC cells, but not their differentiated derivatives, are refractory to the expression of many viral genomes following infection. Based upon these studies, we propose that fundamental changes in gene expression that are observed when differentiation is triggered in EC cells are likely to be due to the disappearance or neutralization of strong repressor elements.

The human pS2 gene, whose expression is restricted to breast cancer cells, and whose transcription is induced by oestrogen in the human breast cancer cell line MCF-7, has been cloned from both placental and MCF-7 cell DNA. The exon-intron organization has been established by electron microscopy using genomic DNA-cDNA or -mRNA hybrid duplexes and by sequencing the exons and exon-intron junctions. The overall organization within and around the pS2 gene is the same in placental and MCF-7 cell DNA and the exonic sequences are identical to those previously determined from the cDNA. The

5'-flanking region of the pS2 gene is also identical (with the exception of two base transitions) in the two tissues. Thus no gene rearrangement nor sequence modification has occurred in the pS2 gene of the malignant and polyploid MCF-7 cells. A TATA-box, a CAAT-box and a GC-rich motif are present in the 5'-flanking region of the pS2 gene, but the latter motif is unusually located between the TATA-box and the capsite. No significant homology could be detected between the 5' flanking sequences of the pS2 gene and those of other oestrogen-responsive genes from different species.

Gene signatures of drug-induced toxicity are of broad interest, but they are often identified from small-scale, single-time point experiments, and are therefore of limited applicability. To address this issue, we performed multivariate analysis of gene expression, cell-based assays, and histopathological data in the TG-GATEs (Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system) database. Data mining highlights four genes-EGR1, ATF3, GDF15 and FGF21-that are induced 2 h after drug administration in human and rat primary hepatocytes poised to eventually undergo cytotoxicity-induced cell death. Modelling and simulation reveals that these early stress-response genes form a functional network with evolutionarily conserved structure and intrinsic dynamics. This is underlined by the fact that early induction of this network in vivo predicts drug-induced liver and kidney pathology with high accuracy. Our findings demonstrate the value of early gene-expression signatures in predicting and understanding compound-induced toxicity. The identified network can empower first-line tests that reduce animal use and costs of safety evaluation.

The Saccharomyces cerevisiae gene deletion collection is widely used for functional gene annotation and genetic interaction analyses. However, the standard G418-resistance cassette used to produce knockout mutants delivers strong regulatory elements into the target genetic loci. To date, its side effects on the expression of neighboring genes have never been systematically assessed. Here, using ribosome profiling data, RT-qPCR, and reporter expression, we investigated perturbations induced by the KanMX module. Our analysis revealed significant alterations in the transcription efficiency of neighboring genes and, more importantly, severe impairment of their mRNA translation, leading to changes in protein abundance. In the 'head-to-head' orientation of the deleted and neighboring genes, knockout often led to a shift of the transcription start site of the latter, introducing new uAUG codon(s) into the expanded 5' untranslated region (5' UTR). In the 'tail-to-tail' arrangement, knockout led to activation of alternative polyadenylation signals in the neighboring gene, thus altering its 3' UTR. These events may explain the so-called neighboring gene effect (NGE), i.e. false genetic interactions of the deleted genes. We estimate that in as much as ∼1/5 of knockout strains the expression of neighboring genes may be substantially (>2-fold) deregulated at the level of translation.

Trypanosoma brucei is an early branching protozoan parasite that causes human and animal African trypanosomiasis. Forward genetics approaches are powerful tools for uncovering novel aspects of trypanosomatid biology, pathogenesis, and therapeutic approaches against trypanosomiasis. Here, we have generated a *T. brucei* cloned ORFeome consisting of >90% of the targeted 7,245 genes and used it to make an inducible gain-of-function parasite library broadly applicable to large-scale forward genetic screens. We conducted a proof-of-principle genetic screen to identify genes whose expression promotes survival in melarsoprol, a critical drug of last resort. The 57 genes identified as overrepresented in melarsoprol survivor populations included the gene encoding the rate-limiting enzyme for the biosynthesis of an established drug target (trypanothione), validating the tool. In addition, novel genes associated with gene expression, flagellum localization, and mitochondrion localization were identified, and a subset of those genes increased melarsoprol resistance upon overexpression in culture. These findings offer new insights into trypanosomatid basic biology, implications for drug targets, and direct or indirect drug resistance mechanisms. This study generated a *T. brucei* ORFeome and gain-of-function parasite library, demonstrated the library's usefulness in forward genetic screening, and identified novel aspects of melarsoprol resistance that will be the subject of future investigations. These powerful genetic tools can be used to broadly advance trypanosomatid research

This article reports on the construction and analysis in vitro and in vivo of novel gene switches that can be used to achieve spatial as well as temporal control over the expression of a transgene of interest. The switches are expected to be functional in virtually any tissue and cell type. They consist of (a) a foreign or modified transactivator expressed under the dual control of a promoter or promoter cassette that is responsive to heat and the transactivator and (b) a promoter responsive to the transactivator for controlling the transgene of interest. A preferred gene switch of this type incorporated a mifepristone-dependent transactivator. This gene switch could be activated by a transient heat treatment in the presence of mifepristone. Activity increased with the intensity of the activating heat treatment and was found to persist for more than 6 days. The gene switch was essentially inactive prior to an activating heat treatment, in the absence or presence of mifepristone. Activated gene switch could be silenced by removal/withdrawal of mifepristone.

Candida dubliniensis is a recently described opportunistic fungal pathogen that is closely related to Candida albicans. Candida dubliniensis readily develops resistance to the azole antifungal agent fluconazole, both in vitro and in infected patients, and this resistance is usually associated with upregulation of the CdMDR1 gene, encoding a multidrug efflux pump of the major facilitator superfamily. To determine the role of CdMDR1 in drug resistance in C. dubliniensis, we constructed an mdr1 null mutant from the fluconazole-resistant clinical isolate CM2, which overexpressed the CdMDR1 gene. Sequential deletion of both CdMDR1 alleles was performed by the MPA(R)-flipping method, which is based on the repeated use of a dominant mycophenolic acid resistance marker for selection of integrative transformants and its subsequent deletion from the genome by FLP-mediated, site-specific recombination. In comparison with its parental strain, the mdr1 mutant showed decreased resistance to fluconazole but not to the related drug ketoconazole. In addition, we found that CdMDR1 confers resistance to the structurally unrelated drugs 4-nitroquinoline-N-oxide, cerulenin, and brefeldin A, since the enhanced resistance to these compounds of the parent strain CM2 compared with the matched susceptible isolate CM1 was abolished in the mdr1 mutant. In contrast, CdMDR1 inactivation did not cause increased susceptibility to amorolfine, terbinafine, fluphenazine, and benomyl, although overexpression of CdMDR1 in a hypersusceptible Saccharomyces cerevisiae strain had previously been shown to confer resistance to these compounds. The effect of CdMDR1 inactivation was identical to that seen in two similarly constructed C. albicans mdr1 mutants. Therefore, despite species-specific differences in the amino acid sequences of the Mdr1 proteins, overexpression of CaMDR1 and CdMDR1 in clinical C. albicans and C. dubliniensis strains seems to confer the same drug resistance profile in both species.

Corynebacterium diphtheriae is a Gram-positive bacterial pathogen and the causative agent of diphtheria, a severe disease of the upper respiratory tract of humans. Factors required for C. diphtheriae to survive in the human host are not well defined, but likely include the acquisition of essential metals such as zinc. In C. diphtheriae, zinc-responsive global gene regulation is controlled by the Zinc Uptake Regulator (Zur), a member of the Fur-family of transcriptional regulators. In this study, we use transcriptomics to identify zinc-regulated genes in C. diphtheriae by comparing gene expression of a wild-type strain grown without and with zinc supplementation. Zur-regulated genes were identified by comparing wild-type gene expression with that of an isogenic zur mutant. We observed zinc repression of several putative surface proteins, the heme efflux system hrtBA, various ABC transporters, and the non-ribosomal peptide synthetase/polyketide synthase cluster sidAB. Furthermore, increased gene expression in response to zinc was observed for the alcohol dehydrogenase, adhA. Zinc and Zur regulation were confirmed for several genes by complementing the zur deletion and subsequent RT-qPCR analysis. We used MEME to predict Zur binding sites within the promoter regions of zinc- and Zur-regulated genes, and verified Zur binding by electrophoretic mobility shift assays. Additionally, we characterized cztA (dip1101), which encodes a putative cobalt/zinc/cadmium efflux family protein. Deletion of cztA results in increased sensitivity to zinc, but not to cobalt or cadmium. This study advances our knowledge of changes to Zur-dependent global gene expression in response to zinc in C.

diphtheriae. The identification of zinc-regulated ABC transporters herein will facilitate future studies to characterize zinc transport in C. diphtheriae.

We report results from the first genome-wide application of a rational drug target selection methodology to a metazoan pathogen genome, the completed draft sequence of Brugia malayi, a parasitic nematode responsible for human lymphatic filariasis. More than 1.5 billion people worldwide are at risk of contracting lymphatic filariasis and onchocerciasis, a related filarial disease. Drug treatments for filariasis have not changed significantly in over 20 years, and with the risk of resistance rising, there is an urgent need for the development of new anti-filarial drug therapies. The recent publication of the draft genomic sequence for B. malayi enables a genome-wide search for new drug targets. However, there is no functional genomics data in B. malayi to guide the selection of potential drug targets. To circumvent this problem, we have utilized the free-living model nematode Caenorhabditis elegans as a surrogate for B. malayi. Sequence comparisons between the two genomes allow us to map C. elegans orthologs to B. malayi genes. Using these orthology mappings and by incorporating the extensive genomic and functional genomic data, including genome-wide RNAi screens, that already exist for C. elegans, we identify potentially essential genes in B. malayi. Further incorporation of human host genome sequence data and a custom algorithm for prioritization enables us to collect and rank nearly 600 drug target candidates. Previously identified potential drug targets cluster near the top of our prioritized list, lending credibility to our methodology. Over-represented Gene Ontology terms, predicted InterPro domains, and RNAi phenotypes of C. elegans orthologs associated with the potential target pool are identified. By virtue of the selection procedure, the potential B. malayi drug targets highlight components of key processes in nematode biology such as central metabolism, molting and regulation of gene expression.

Data-driven discovery of cancer driver genes, including tumor suppressor genes (TSGs) and oncogenes (OGs), is imperative for cancer prevention, diagnosis, and treatment. Although epigenetic alterations are important for tumor initiation and progression, most known driver genes were identified based on genetic alterations alone. Here, we developed an algorithm, DORGE (Discovery of Oncogenes and tumor suppressoR genes using Genetic and Epigenetic features), to identify TSGs and OGs by integrating comprehensive genetic and epigenetic data. DORGE identified histone modifications as strong predictors for TSGs, and it found missense mutations, super enhancers, and methylation differences as strong predictors for OGs. We extensively validated DORGE-predicted cancer driver genes using independent functional genomics data. We also found that DORGE-predicted dual-functional genes (both TSGs and OGs) are enriched at hubs in protein-protein interaction and drug-gene networks. Overall, our study has deepened the understanding of epigenetic mechanisms in tumorigenesis and revealed previously undetected cancer driver genes

It is found that the frequencies of the selected genes follow a power-law distribution, indicating that only a few top-ranked genes can be used as potential diagnosis biomarkers. Moreover, the top-ranked genes leading to very high prediction accuracy are closely related to specific tumor subtype and even hub genes. Compared with other related methods, the proposed method can achieve higher prediction accuracy with fewer genes. Moreover, they are further justified by analyzing the top-ranked genes in the context of individual gene function, biological pathway, and protein-protein interaction network.

Present study produces for the first time, a signature, in the form of a robust list of gene circuitry whose presence or absence could potentially define the pathogenicity of a microbiome. Extensive literature search substantiated a bulk majority of the commensal and pathogenic circuitry in our predicted list. Scanning microbiome libraries for these circuitry motifs will provide further insights into the complex and context dependent pathogenicity of bacteria.

In primary cultures of adult rat hepatocytes, transcription of the albumin gene, measured as incorporation of [alpha-32P]UTP into mRNA in isolated nuclei, decreased dramatically during culture without addition of serum and hormone, becoming almost negligible 10 h after plating. Of the hormones

tested, dexamethasone (0.1 microM) prevented this decrease and restored the transcription within 2 h to the same level as that before culture. The half-maximum dose of dexamethasone for induction of transcription of the albumin gene was about 30 nM. The in vitro finding that expression of the albumin gene is strictly regulated by glucocorticoid was confirmed by an in vivo experiment in adrenalectomized rats showing that the transcription decreased markedly 14 days after adrenalectomy, but was restored rapidly by administration of hydrocortisone. This finding was also supported by identification of a glucocorticoid regulatory sequence from -50 to -62 base pairs between the TATA box and CAT box upstream of the 5'-end of the albumin gene. Cycloheximide inhibited the induction of transcription of the albumin gene by dexamethasone, suggesting that a rapidly induced mediator protein, which is also regulated by glucocorticoid, is involved in the induction of albumin gene expression by glucocorticoid. The albumin gene was also regulated by various other hormones besides glucocorticoid. Glucagon markedly enhanced the transcription induced by dexamethasone, although glucagon alone had no effect. Conversely, epinephrine suppressed stimulation of expression of the albumin gene by dexamethasone. Insulin and triiodothyronine had no effect on transcription of the albumin gene. From these findings we conclude that expression of the albumin gene depends strictly on glucocorticoid, and this dependence is modulated by other hormones.

Thiopeptides are a growing class of sulfur-rich, highly modified heterocyclic peptides that are mainly active against Gram-positive bacteria including various drug-resistant pathogens. Recent studies also reveal that many thiopeptides inhibit the proliferation of human cancer cells, further expanding their application potentials for clinical use. Thiopeptide biosynthesis shares a common paradigm, featuring a ribosomally synthesized precursor peptide and conserved posttranslational modifications, to afford a characteristic core system, but differs in tailoring to furnish individual members. Identification of new thiopeptide gene clusters, by taking advantage of increasing information of DNA sequences from bacteria, may facilitate new thiopeptide discovery and enrichment of the unique biosynthetic elements to produce novel drug leads by applying the principle of combinatorial biosynthesis. In this study, we have developed a web-based tool ThioFinder to rapidly identify thiopeptide biosynthetic gene cluster from DNA sequence using a profile Hidden Markov Model approach. Fifty-four new putative thiopeptide biosynthetic gene clusters were found in the sequenced bacterial genomes of previously unknown producing microorganisms. ThioFinder is fully supported by an open-access database ThioBase, which contains the sufficient information of the 99 known thiopeptides regarding the chemical structure, biological activity, producing organism, and biosynthetic gene (cluster) along with the associated genome if available. The ThioFinder website offers researchers a unique resource and great flexibility for sequence analysis of thiopeptide biosynthetic gene clusters.

Bud outgrowth is under the intricate control of environmental and endogenous factors. In a recent paper, [1] we demonstrated that light perceived by Rosa buds triggers cytokinins (CK) synthesis within 3 hours in the adjacent node followed by their transport to the bud. There, CK control expression of a set of major genes (strigolactones-, auxin-, sugar sink strength-, cells division and elongation-related genes) leading to bud outgrowth in light. Conversely, under dark condition, CK accumulation and transport to the bud are repressed and no bud outgrowth occurs. In this paper, we show that the 3 expansin genes RhEXPA1,2,3 are under the control of both light and CK during bud outgrowth. In silico analysis of promoter sequences highlights 2 regions enriched in light and CK cis-regulatory elements as well as a specific cis-element in pRhEXPA3, potentially responsible for the expression patterns observed in response to CK and light.

In Escherichia coli, there are 32 open reading frames (ORFs) that are assumed to be response regulator genes of two-component signal transduction systems on the basis of sequence similarities. We cloned all of these 32 ORFs into a multicopy expression vector and investigated whether or not they confer drug resistance via control of drug resistance determinants. Fifteen of these ORFs, i.e., baeR, citB, cpxR, evgA, fimZ, kdpE, narL, narP, ompR, rcsB, rstA, torR, yedW, yehT, and dcuR, conferred increased single- or multidrug resistance. Two-thirds of them conferred deoxycholate resistance. Five of them, i.e., evgA, baeR, ompR, cpxR, and rcsB, modulated the expression of several drug exporter

genes. The drug resistance mediated by evgA, baeR, and cpxR could be assigned to drug exporters by using drug exporter gene knockout strains.

The DNA element governing the inducible expression of drug-metabolizing P-450c gene by xenobiotic treatments was investigated by gene transfer methods. A variety of dissected fragments from -844 to -1140bp region which was essential for the inducibility of P-450c gene were placed on the heterologous SV40 promoter for testing the inducibility. Mapping studies in combination with gel retardation assay defined the presence of the two xenobiotic responsive elements (XRE, XRE1, -1007 - -1021bp; XRE2, -1088 - -1092bp) composed of about 15 nucleotides which expressed the enhancer activity in response to xenobiotic inducers. The two XREs share 10 nucleotides in common out of 15 as expressed in the sequence CG/CTG/CC/TTG/CTCACGCT/AA and are arranged in the inverse orientation. They are different from DREs (drug responsive element) proposed previously (Sogawa, K. et al. Proc. Natl. Acad. Sci. 83, 8044-8048 (1986] and expressed a strong enhancer activity in response to 3-methylcholanthrene. The XRE shows a significant homology with glucocorticoid regulatory elements and apparently needs normal functions of a putative xenobiotic receptor for the inducible enhancer activity.

Previously, we had established a highly sensitive trap vector system for the efficient isolation of reporter cells for a certain condition of interest. In this study, we used this system to screen reporter cells that express the luciferase and enhanced green fluorescent protein genes in response to dexamethasone, a glucocorticoid receptor agonist to facilitate glucocorticoid signaling research. In total, 10 clones were isolated. The insertion sites of the trap vector were analyzed using 5' rapid amplification of cDNA ends (5' RACE), whereupon LPIN1, PKP2, and FKBP5 were identified as genes that were upregulated by the dexamethasone treatment. Specifically, PKP2 has not previously been focused as a gene that responds to glucocorticoids. The PKP2 mRNA was analyzed and induction of the endogenous gene was confirmed by real-time polymerase chain reaction. Given that PKP2 does not appear to have a consensus glucocorticoid response element (GRE) sequence, this reporter clone could supplement the current GRE-based reporter systems that are prevalently used. Because different clones showed different responses to glucocorticoids, these clones should provide more information than analysis with a single reporter clone. This paper demonstrates that the previously developed trap vector technology can contribute to the rapid construction of drug evaluation systems.

Chemical insecticides have been heavily employed as the most effective measure for control of agricultural and medical pests, but evolution of resistance by pests threatens the sustainability of this approach. Resistance-conferring mutations sometimes impose fitness costs, which may drive subsequent evolution of compensatory modifier mutations alleviating the costs of resistance. However, how modifier mutations evolve and function to overcome the fitness cost of resistance still remains unknown. Here we show that overexpression of P450s not only confers imidacloprid resistance in the brown planthopper, Nilaparvata lugens, the most voracious pest of rice, but also leads to elevated production of reactive oxygen species (ROS) through metabolism of imidacloprid and host plant compounds. The inevitable production of ROS incurs a fitness cost to the pest, which drives the increase or fixation of the compensatory modifier allele T65549 within the promoter region of N. lugens peroxiredoxin (NlPrx) in the pest populations. T65549 allele in turn upregulates the expression of NlPrx and thus increases resistant individuals' ability to clear the cost-incurring ROS of any source. The frequent involvement of P450s in insecticide resistance and their capacity to produce ROS while metabolizing their substrates suggest that peroxiredoxin or other ROS-scavenging genes may be among the common modifier genes for alleviating the fitness cost of insecticide resistance.

A copper resistance gene cluster (6 genes, ~8.2 kb) was isolated from the cyanobacterium Synechocystis sp. PCC 6803 by recombineering recovery (RR). Following integration of a narrow-host-range plasmid vector adjacent to the target region in the Synechocystis genome (pSYSX), DNA was isolated from transformed cells and the plasmid plus flanking sequence circularized by recombineering to precisely clone the gene cluster. Complementation of a copper-sensitive Escherichia coli mutant

demonstrated the functionality of the pcopM gene encoding a copper-binding protein. RR provides a novel alternative method for cloning large DNA fragments from species that can be transformed by homologous recombination.

The past three years can be considered in cardiology as critical for understanding the relevance of developmental genes in the adult cardiac physiology. Also, for the first time, endogenous control of programmed cell death has been demonstrated to mark the transition between normal adaptation and cardiac hypertrophy. Most of this work has been based on previous analysis using molecular markers of cardiac determination and differentiation, work that has served a double aim: First, the determination of the cellular process that contribute to the specification of the working heart and secondly, the characterization of key regulatory factors in cardiogenesis. These studies in conjunction with the recent availability of single gene mutation in transgenic mice have furnished a new perspective in the nature of cardiac defects either in shape or function. Here we review some of the key factors in cardiac morphogenesis from the perspective of the analysis of gene mutation.

Diseases caused by various Mycobacterium sp., especially Mycobacterium tuberculosis, are a major burden on global health care. Due to high intrinsic antibiotic resistance, treatment options are severely limited. In mycobacteria, WhiB7 coordinates intrinsic resistance to a broad range of antibiotics. While WhiB7 has been established as an auto-regulatory transcriptional activator, the signals and genes needed to induce its expression are poorly understood. Using Mycobacterium smegmatis as a model, we coupled transposon mutagenesis and next generation sequencing with WhiB7-specific antibiotic selection to identify genes that contribute to WhiB7 regulation and function. We showed that whiB7 expression was regulated by two coordinated processes: early termination of the whiB7 transcript and increased whiB7 promoter activity. Early termination was irreversibly maintained by constitutive expression of a putative aspartate aminotransferase gene, MSMEG_4060. A pair of hypothetical genes, MSMEG_3637 and MSMEG_3638, were identified as important contributors to whiB7 promoter induction on antibiotic challenge. Expansion of our understanding of the WhiB7-resistance pathway may lead to identification of inhibitors that allow the use of previously ineffective antibiotics to treat mycobacterial diseases.

While the fundamental building blocks of biology are being tabulated by the various genome projects, microarray technology is setting the stage for the task of deducing the connectivity of large-scale gene networks. We show how the perturbation of carefully chosen genes in a microarray experiment can be used in conjunction with a reverse engineering algorithm to reveal the architecture of an underlying gene regulatory network. Our iterative scheme identifies the network topology by analyzing the steady-state changes in gene expression resulting from the systematic perturbation of a particular node in the network. We highlight the validity of our reverse engineering approach through the successful deduction of the topology of a linear in numero gene network and a recently reported model for the segmentation polarity network in Drosophila melanogaster. Our method may prove useful in identifying and validating specific drug targets and in deconvolving the effects of chemical compounds.

DOK is a new type of regulatory protein family that participates in the regulation of tumor cell growth. However, most of the studies are conducted in cell lines, and systematic studies have not been conducted in human tumors.

Cone snails, which are predatory marine gastropods, produce a cocktail of venoms used for predation, defense and competition. The major venom component, conotoxin, has received significant attention because it is useful in neuroscience research, drug development and molecular diversity studies. In this study, we report the genomic characterization of nine conotoxin gene superfamilies from 18 Conus species and investigate the relationships among conotoxin gene structure, molecular evolution and diversity. The I1, I2, M, O2, O3, P, S, and T superfamily precursors all contain three exons and two introns, while A superfamily members contain two exons and one intron. The introns are conserved within a certain gene superfamily, and also conserved across different Conus species, but divergent

among different superfamilies. The intronic sequences contain many simple repeat sequences and regulatory elements that may influence conotoxin gene expression. Furthermore, due to the unique gene structure of conotoxins, the base substitution rates and the number of positively selected sites vary greatly among exons. Many more point mutations and trinucleotide indels were observed in the mature peptide exon than in the other exons. In addition, the first example of alternative splicing in conotoxin genes was found. These results suggest that the diversity of conotoxin genes has been shaped by point mutations and indels, as well as rare gene recombination or alternative splicing events, and that the unique gene structures could have made a contribution to the evolution of conotoxin genes.

Of 173 mutants of Saccharomyces cerevisiae resistant to the antimitotic drug benomyl (BenR), six also conferred cold-sensitivity for growth and three others conferred temperature-sensitivity for growth in the absence of benomyl. All of the benR mutations tested, including the nine conditional-lethal mutations, were shown to be in the same gene. This gene, TUB2, has previously been molecularly cloned and identified as the yeast structural gene encoding beta-tubulin. Four of the conditional-lethal alleles of TUB2 were mapped to particular restriction fragments within the gene. One of these mutations was cloned and sequenced, revealing a single amino acid change, from arginine to histidine at amino acid position 241, which is responsible for both the BenR and the cold-sensitive lethal phenotypes. The terminal arrest morphology of conditional-lethal alleles of TUB2 at their restrictive temperature showed a characteristic cell-division-cycle defect, suggesting a requirement for tubulin function primarily in mitosis during the vegetative growth cycle. The TUB2 gene was genetically mapped to the distal left arm of chromosome VI, very near the actin gene, ACT1; no CDC (cell-division-cycle) loci have been mapped previously to this location. TUB2 is thus the first cell-division-cycle gene known to encode a cytoskeletal protein that has been identified in S. cerevisiae.

A fish calmodulin (CaM) gene was characterized for the first time in grass carp. The CaM gene is about 12-Kb in size with identical intron/exon organization as that of mammalian CaM genes. When compared to mammalian counterparts, the 5'-promoter region of grass carp CaM gene contains a TATA box and has a much lower GC content and CpG dinucleotide frequency. Interestingly, the 5'-promoter of carp CaM gene is AT-rich with multiple IRS elements and putative binding sites for Pit-1, Sp1/Sp3 and AP1. Using luciferase reporter assay, a potent silencer region was identified in the distal region of grass carp CaM promoter. Besides, the CaM promoter activity could be upregulated by IGF but suppressed by PACAP, forskolin and over-expression of Sp1 and Sp3. These findings, taken together, indicate that grass carp CaM gene does not exhibit the typical features of housekeeping genes and its expression is under the control of hormone factors, presumably by coupling with the appropriate signaling pathways/transcription factors.

The cumulative of genes carrying mutations is vital for the establishment and development of cancer. However, this driver gene exploring research line has selected and used types of tools and models of analysis unsystematically and discretely. Also, the previous studies may have neglected low-frequency drivers and seldom predicted subgroup specificities of identified driver genes. In this study, we presented an improved driver gene identification and analysis pipeline that comprises the four most widely focused analyses for driver genes: enrichment analysis, clinical feature association with expression profiles of identified driver genes as well as with their functional modules, and patient stratification by existing advanced computational tools integrating multi-omics data. The improved pipeline's general usability was demonstrated straightforwardly for breast cancer, validated by some independent databases. Accordingly, 31 validated driver genes, including four novel ones, were discovered. Subsequently, we detected cancer-related significantly enriched gene ontology terms and pathways, probable drug targets, two co-expressed modules associated significantly with several clinical features, such as number of positive lymph nodes, Nottingham prognostic index, and tumor stage, and two biologically distinct groups of BRCA patients.

Proto-oncogenes are cellular genes that are expressed during normal growth and developmental processes. Altered versions of normal proto-oncogenes have been implicated in the development of

human neoplasia. In this report, we show the detection of activated proto-oncogenes in various spontaneous and chemically induced rodent tumors. The majority of activated proto-oncogenes found in these tumors are members of the ras gene family and have been activated by a point mutation. Characterization of the activating mutation may be useful in determining whether this proto-oncogene was activated by direct interaction of the chemical with the DNA. Comparison of activating lesions in spontaneous versus chemically induced tumors should be helpful in determining whether the chemical acts via a genotoxic or a nongenotoxic mechanism. All of this information may be helpful in the assessment of potential carcinogenic hazards of human exposure to chemicals.

DNA methylation and demethylation precisely and effectively modulate gene expression during plant growth and development and in response to stress. However, expression profiles of genes involved in DNA methylation and demethylation during plant development and their responses to phytohormone treatments remain largely unknown. We characterized the spatiotemporal expression patterns of genes involved in de novo methylation, methyl maintenance, and active demethylation in roots, shoots, and reproductive organs using β-glucuronidase (GUS) reporter lines. Promoters of DNA demethylases were generally more highly active at the mature root tissues, whereas the promoters of genes involved in DNA methylation were more highly active at fast-growing root tissues. The promoter activity also implies that methylation status in shoot apex, leaf primordia, floral organs, and developing embryos is under tight equilibrium through the activity of genes involved in DNA methylation and demethylation. The promoter activity of DNA methylation and demethylation-related genes in response to various phytohormone treatments revealed that phytohormones can alter DNA methylation status in specific and redundant ways. Overall, our results illustrate that DNA methylation and demethylation pathways act synergistically and antagonistically in various tissues and in response to phytohormone treatments and point to the existence of hormone-linked methylome regulation mechanisms that may contribute to tissue differentiation and development.

Bacterial pathogens evolve during the course of infection as they adapt to the selective pressures that confront them inside the host. Identification of adaptive mutations and their contributions to pathogen fitness remains a central challenge. Although mutations can either target intergenic or coding regions in the pathogen genome, studies of host adaptation have focused predominantly on molecular evolution within coding regions, whereas the role of intergenic mutations remains unclear. Here, we address this issue and investigate the extent to which intergenic mutations contribute to the evolutionary response of a clinically important bacterial pathogen, Pseudomonas aeruginosa, to the host environment, and whether intergenic mutations have distinct roles in host adaptation. We characterize intergenic evolution in 44 clonal lineages of P. aeruginosa and identify 77 intergenic regions in which parallel evolution occurs. At the genetic level, we find that mutations in regions under selection are located primarily within regulatory elements upstream of transcriptional start sites. At the functional level, we show that some of these mutations both increase or decrease transcription of genes and are directly responsible for evolution of important pathogenic phenotypes including antibiotic sensitivity. Importantly, we find that intergenic mutations facilitate essential genes to become targets of evolution. In summary, our results highlight the evolutionary significance of intergenic mutations in creating host-adapted strains, and that intergenic and coding regions have different qualitative contributions to this process.

Tyrosine hydroxylase is the rate-limiting enzyme in the biosynthesis of catecholamines. Expression of the tyrosine hydroxylase gene is regulated at the transcriptional level by extracellular signalling molecules, including epidermal growth factor (EGF), nerve growth factor (NGF) and glucocorticoids. We have analysed the stimulation of tyrosine hydroxylase gene transcription by the phorbol ester 12-O-tetradecanoylphorbol-13-acetate (TPA) in noradrenergic locus coeruleus-like CATH.a cells and observed a striking enhancement of the transcriptional activation potential of the ternary complex factor Ets-like protein-1 (Elk-1), a key transcriptional regulator of serum response element-driven gene transcription. Likewise, TPA strongly up-regulated the biosynthesis of the transcription factor Egr-1 via distal serum response elements within the Egr-1 5'-flanking region. Subsequently, enhancement of the

transcriptional activation potential of Egr-1 was observed. Overexpression of Egr-1 was sufficient to activate transcription of a tyrosine hydroxylase promoter/reporter gene, corroborating the view that the tyrosine hydroxylase gene is a target gene of Egr-1. Expression of dominant-negative mutants of Elk-1 or Egr-1 impaired TPA-induced stimulation of a tyrosine hydroxylase promoter/reporter gene transcription. In contrast, dominant-negative mutants of the transcription factors activating transcription factor (ATF)-2, ATF4, cAMP response element-binding protein, c-Jun and CCAAT/enhancer binding protein (C/EBP) did not change TPA-induced tyrosine hydroxylase promoter activity, indicating that these proteins are not part of the TPA-mediated signalling cascade directed towards the tyrosine hydroxylase gene.

Two genes, MDR1 and MDR3, constitute the human P-glycoprotein gene family. To examine the evolutionary relationship between the three known classes of mammalian P-glycoprotein genes, we have cloned the MDR3 gene and compared its structure with that of the human MDR1 and the mouse mdr1 (mdr1b) genes analyzed by other groups. The MDR3 gene contains 28 exons and 27 of these contain coding sequences for the two homologous halves of the protein that correlate with functional domains. This structure is virtually identical to that of the human MDR1 gene and the mouse mdr1 (mdr1b) gene, indicating that the exon/intron structure was fixed before the duplication events that generated different classes of P-glycoproteins, but after the P-glycoproteins diverged from related genes, like the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which has an entirely different exon/intron structure. The four alternatively spliced transcripts of the MDR3 gene arise from alternative splicing of exons 23 and 26. Our analysis of DNA clones covering about 120 kilobases (kb) of the human MDR locus, including the entire MDR3 gene (74 kb) and the intergenic region between both genes (34 kb), combined with pulsed-field gel electrophoresis data shows that the human MDR locus covers about 230 kb. In contrast to the mouse mdr genes, both human genes are transcribed in the same direction (MDR3 located downstream of MDR1). The CpG-rich sequences marking the 5' ends of both genes are hypomethylated to different extents in different cell lines. Hypomethylation roughly correlates with transcriptional activity.

Reverse transcription (RT)-PCR is a valuable tool widely used for analysis of gene expression. In bacteria, RT-PCR is helpful beyond standard protocols of northern blot RNA/DNA hybridization (to identify transcripts) and primer extension (to locate their start points), as these methods have been difficult with transcripts that are low in abundance or unstable, similar to long multi-gene operons. In this report, RT-PCR is adapted to analyze transcripts that form long multi-gene operons--where they start and where they stop. The transcripts can also be semiquantitated to follow the expression of genes under different growth conditions. Examples using RT-PCR are presented with two different multi-gene systems for metal cation resistance to silver and mercury ions. The silver resistance system [9 open reading frames (ORFs); 12.5 kb] is shown by RT-PCR to synthesize three nonoverlapping messenger RNAs that are transcribed divergently. In the mercury resistance system (8 ORFs; 6.3 kb), all the genes are transcribed in the same orientation, and two promoter sites produce overlapping transcripts. For RT-PCR, reverse transcriptase enzyme is used to synthesize first-strand cDNA that is used as a template for PCR amplification of single-gene products, from the beginning, middle or end of long multi-gene, multi-transcript gene clusters.

Glucocorticoid and cyclic AMP increase tyrosine hydroxylase (TH) activity and mRNA levels in pheochromocytoma cultures. The transcriptional activity of the TH gene, as measured by nuclear run-on assay, is also increased when cultures are treated with the synthetic glucocorticoid dexamethasone or agents that increase intracellular cyclic AMP, such as forskolin and 8-BrcAMP. Both inducers effect transcriptional changes within 10 min after treatment and are maximal after 30 min for forskolin and after 60 min for dexamethasone. The 5' flanking sequences of the TH gene were fused to the bacterial gene chloramphenicol acetyltransferase (CAT), and the hybrid gene was transfected into pheochromocytoma cultures and GH4 pituitary cells. In both cell lines, a region of the TH gene containing bases -272 to +27 conferred induction of CAT by cyclic AMP, but not by glucocorticoid. The same results were found when a region of the TH gene containing -773 to +27 was used. Thus, the

sequences required for induction of TH by cyclic AMP are contained within 272 bases of 5' flanking sequence, but sequences sufficient for glucocorticoid regulation are not contained within 773 bases.

The molecular mechanisms by which plants acclimate to oxidative stress are poorly understood. To identify the processes involved in acclimation, we performed a comprehensive analysis of gene expression in Nicotiana tabacum leaves acclimated to oxidative stress. Combining mRNA differential display and cDNA array analysis, we estimated that at least 95 genes alter their expression in tobacco leaves acclimated to oxidative stress, of which 83% are induced and 17% repressed. Sequence analysis of 53 sequence tags revealed that, in addition to antioxidant genes, genes implicated in abiotic and biotic stress defenses, cellular protection and detoxification, energy and carbohydrate metabolism, de novo protein synthesis, and signal transduction showed altered expression. Expression of most of the genes was enhanced, except for genes associated with photosynthesis and light-regulated processes that were repressed. During acclimation, two distinct groups of coregulated genes ("early-" and "late-response" gene regulons) were observed, indicating the presence of at least two different gene induction pathways. These two gene regulons also showed differential expression patterns on an oxidative stress challenge. Expression of "late-response" genes was augmented in the acclimated leaf tissues, whereas expression of "early-response" genes was not. Together, our data suggest that acclimation to oxidative stress is a highly complex process associated with broad gene expression adjustments. Moreover, our data indicate that in addition to defense gene induction, sensitization of plants for potentiated gene expression might be an important factor in oxidative stress acclimation.

Functionally interacting perturbations, such as synergistic drugs pairs or synthetic lethal gene pairs, are of key interest in both pharmacology and functional genomics. However, to find such pairs by traditional screening methods is both time consuming and costly. We present a novel computational-experimental framework for efficient identification of synergistic target pairs, applicable for screening of systems with sizes on the order of current drug, small RNA or SGA (Synthetic Genetic Array) libraries (>1000 targets). This framework exploits the fact that the response of a drug pair in a given system, or a pair of genes' propensity to interact functionally, can be partly predicted by computational means from (i) a small set of experimentally determined target pairs, and (ii) pre-existing data (e.g. gene ontology, PPI) on the similarities between targets. Predictions are obtained by a novel matrix algebraic technique, based on cyclical projections onto convex sets. We demonstrate the efficiency of the proposed method using drug-drug interaction data from seven cancer cell lines and gene-gene interaction data from yeast SGA screens. Our protocol increases the rate of synergism discovery significantly over traditional screening, by up to 7-fold. Our method is easy to implement and could be applied to accelerate pair screening for both animal and microbial systems.

Identification of novel targets for the development of more effective antimalarial drugs and vaccines is a primary goal of the Plasmodium genome project. However, deciding which gene products are ideal drug/vaccine targets remains a difficult task. Currently, a systematic disruption of every single gene in Plasmodium is technically challenging. Hence, we have developed a computational approach to prioritize potential targets. A pathway/genome database (PGDB) integrates pathway information with information about the complete genome of an organism. We have constructed PlasmoCyc, a PGDB for Plasmodium falciparum 3D7, using its annotated genomic sequence. In addition to the annotations provided in the genome database, we add 956 additional annotations to proteins annotated as "hypothetical" using the GeneQuiz annotation system. We apply a novel computational algorithm to PlasmoCyc to identify 216 "chokepoint enzymes." All three clinically validated drug targets are chokepoint enzymes. A total of 87.5% of proposed drug targets with biological evidence in the literature are chokepoint reactions. Therefore, identifying chokepoint enzymes represents one systematic way to identify potential metabolic drug targets.

Schistosome worms of the genus Schistosoma are the causative agents of schistosomiasis, a devastating parasitic disease affecting more than 240 million people worldwide. Schistosomes have complex life cycles, and have been challenging to manipulate genetically due to the dearth of molecular tools.

Although the use of gene overexpression, gene knockouts or knockdowns are straight-forward genetic tools applied in many model systems, gene misexpression and genetic manipulation of schistosome genes in vivo has been exceptionally challenging, and plasmid based transfection inducing gene expression is limited. We recently reported the use of polyethyleneimine (PEI) as a simple and effective method for schistosome transfection and gene expression. Here, we use PEI-mediated schistosome plasmid transgenesis to define and compare gene expression profiles from endogenous and nonendogenous promoters in the schistosomula stage of schistosomes that are potentially useful to misexpress (underexpress or overexpress) gene product levels. In addition, we overexpress schistosome genes in vivo using a strong promoter and show plasmid-based misregulation of genes in schistosomes, producing a clear and distinct phenotype--death. These data focus on the schistosomula stage, but they foreshadow strong potential for genetic characterization of schistosome molecular pathways, and potential for use in overexpression screens and drug resistance studies in schistosomes using plasmid-based gene expression.

Loss of heterozygosity (LOH) at human chromosome 18q, which includes the gene Deleted in Colorectal Cancer (DCC), has been linked to colorectal and many other human cancers. DCC encodes the receptor for the axon guidance molecule Netrin (Net) and functions during neural development in a variety of organisms. However, since its discovery in the 1990s, the status of DCC as a tumor suppressor has been debated, primarily due to a lack of support for this hypothesis in animal models. A recent study from our laboratory capitalized on the genetic tractability of Drosophila melanogaster to demonstrate that this gene functions as an invasive tumor suppressor, thereby providing the first direct link between DCC loss and metastatic phenotypes in an animal model for cancer. Two subsequent studies from other laboratories have demonstrated that DCC suppresses tumor progression and metastasis in murine colorectal and mammary tumor models. Combined, these findings have prompted the rebirth of DCC as a tumor suppressor and highlighted the need for continued analysis of DCC function in animal models for human cancer.

The study of the minimum set of genes required to sustain life is a fundamental question in biological research. Recent studies on bacterial essential genes suggested that between 350 and 700 genes are essential to support autonomous bacterial cell growth. Essential genes are of interest as potential new antimicrobial drug targets; hence, our aim was to identify the essential genome of the cystic fibrosis (CF) isolate *Burkholderia cenocepacia* H111. Using a transposon sequencing (Tn-Seq) approach, we identified essential genes required for growth in rich medium under aerobic and microoxic conditions as well as in a defined minimal medium with citrate as a sole carbon source. Our analysis suggests that 398 genes are required for autonomous growth in rich medium, a number that represents only around 5% of the predicted genes of this bacterium. Five hundred twenty-six genes were required to support growth in minimal medium, and 434 genes were essential under microoxic conditions (0.5% O_2). A comparison of these data sets identified 339 genes that represent the minimal set of essential genes required for growth under all conditions tested and can be considered the core essential genome of *B. cenocepacia* H111. The majority of essential genes were found to be located on chromosome 1, and few such genes were located on chromosome 2, where most of them were clustered in one region. This gene cluster is fully conserved in all *Burkholderia* species but is present on chromosome 1 in members of the closely related genus *Ralstonia*, suggesting that the transfer of these essential genes to chromosome 2 in a common ancestor contributed toward the separation of the two genera.

Gene trap mutagenesis in mouse embryonic stem cells has been widely used for genome-wide studies of mammalian gene function. However, while large numbers of genes can be disrupted, individual mutations may suffer from limitations due to the structure and/or placement of targeting vector. To extend the utility of gene trap mutagenesis, replaceable 3' [or poly(A)] gene trap vectors were developed that permit sequences inserted in individual entrapment clones to be engineered by Cre-mediated recombination. 3' traps incorporating different drug resistance genes could be readily exchanged, simply by selecting for the drug-resistance gene of the replacement vector. By substituting different 3' traps, we show that otherwise identical fusion genes containing a large first exon (804 nt) are not expressed

at appreciably lower levels than genes expressing small first exons (384 and 151 nt). Thus, size appears to have less effect on the expression and processing of first exons than has been reported for internal exons. Finally, a retroviral poly(A) trap (consisting of a RNA polymerase II promoter, a neomycin-resistance gene, and 5'-splice site) typically produced mutagenized clones in which vector sequences spliced to the 3'-terminal exons of cellular transcription units, suggesting strong selection for fusion transcripts that evade nonsense-mediated decay. The efficient exchange of poly(A) traps should greatly extend the utility of mutant libraries generated by gene entrapment and provides new strategies to study the rules that govern the expression of exons inserted throughout the genome.

Understanding the mechanisms of tissue-specific transcriptional regulation is crucial as mis-regulation can cause a broad range of diseases. Here, we investigated transcription factors (TF) that are indispensable for the topological control of tissue specific and cell-type specific regulatory networks as a function of their binding to regulatory elements on promoters and enhancers of corresponding target genes. In particular, we found that promoter-binding TFs that were indispensable for regulatory network control regulate genes that are tissue-specifically expressed and overexpressed in corresponding cancer types. In turn, indispensable, enhancer-binding TFs were enriched with disease and signaling genes as they control an increasing number of cell-type specific regulatory networks. Their target genes were cell-type specific for blood and immune-related cell-types and over-expressed in blood-related cancers. Notably, target genes of indispensable enhancer-binding TFs in cell-type specific regulatory networks were enriched with cancer drug targets, while target genes of indispensable promoter-binding TFs were bona-fide targets of cancer drugs in corresponding tissues. Our results emphasize the significant role control analysis of regulatory networks plays in our understanding of transcriptional regulation, demonstrating potential therapeutic implications in tissue-specific drug discovery research.

The insulin-like growth factor (IGF-1) signalling is highly implicated in cancer. In this signalling the IGF-1 receptor (IGF-1R) is unquestionable, the predominating single factor. IGF-1R is crucial for tumour transformation and survival of malignant cell, but is only partially involved in normal cell growth. This is in part due to the interactions with oncogenes. Recent findings suggest a close interplay with the p53/MDM2 pathway. Disturbances in components in the p53/MDM2/IGF-1R network may cause IGF-1R upregulation and growth advantage for the cancer cell. Targeting of IGF-1R is more and more seen as a promising option for future cancer therapy. Single chain antibodies and small molecules with selective effects on IGF-1R dependent malignant growth are of particular interest. Forthcoming clinical trials are welcome and will indeed be the only way to evaluate the impact of IGF-1R targeting in human cancer.

A cosmid clone bank of yeast DNA has been used to isolate the cycloheximide resistance gene cyh2 of Saccharomyces cerevisiae. A cosmid carrying this gene was identified by cross hybridization to another cloned gene, tsm437. The two genes, which are tightly linked genetically are both present on a 31 kb segment of cloned DNA. The cyh2 gene encodes ribosomal protein L29, a component of the large subunit. Blot hybridization analysis reveals that this gene is present as a single copy in the yeast genome, unlike many other yeast ribosomal protein genes which appear to be duplicated. The cyh2 gene also appears to contain an intervening sequence, a characteristic common to most yeast ribosomal protein genes that have been cloned.

The human bcl-2 gene contains a GC-rich region upstream of the P1 promoter that has been shown to be critically involved in the regulation of bcl-2 gene expression. We have demonstrated that the guanine-rich strand of the DNA in this region can form any one of three distinct intramolecular G-quadruplex structures. Mutation and deletion analysis permitted isolation and identification of three overlapping DNA sequences within this element that formed the three individual G-quadruplexes. Each of these was characterized using nondenaturing gel analysis, DMS footprinting, and circular dichroism. The central G-quadruplex, which is the most stable, forms a mixed parallel/antiparallel structure consisting of three tetrads connected by loops of one, seven, and three bases. Three different G-quadruplex-interactive agents were found to further stabilize these structures, with individual selectivity

toward one or more of these G-quadruplexes. Collectively, these results suggest that the multiple G-quadruplexes identified in the promoter region of the bcl-2 gene are likely to play a similar role to the G-quadruplexes in the c-myc promoter in that their formation could serve to modulate gene transcription. Last, we demonstrate that the complexity of the G-quadruplexes in the bcl-2 promoter extends beyond the ability to form any one of three separate G-quadruplexes to each having the capacity to form either three or six different loop isomers. These results are discussed in relation to the biological significance of this G-quadruplex-forming element in modulation of bcl-2 gene expression and the inherent complexity of the system where different G-quadruplexes and loop isomers are possible.

The gene encoding the manganese-containing superoxide dismutase (SOD) of Halobacterium cutirubrum was isolated and characterized. The gene and 5'- and 3'-untranslated regions were located on a genomic DNA fragment of 1127 nucleotides. The deduced amino acid sequence is 200 residues long and has 39-42% identity with manganese-containing SODs of eubacteria and mitochondria. This homology may be due to either lateral transfer of the gene between eubacteria and archaebacteria or to high amino acid sequence conservation in the enzyme during the separate evolution of eubacteria and archaebacteria. Transcription of the gene initiates only about three nucleotides upstream of the translation initiation codon. The 5' end of the transcript does not contain a purine-rich Shine-Dalgarno sequence, and the promoter region does not contain consensus sequences found in other archaebacterial promoters. Termination of transcription occurs at 5 consecutive thymine residues that are preceded by a GC-rich region. The gene is basally expressed in anaerobically grown cells but is also inducible by paraquat, a generator of oxygen radicals. The same transcription initiation site is used in both types of expression, suggesting that one promoter is responsible for both basal and regulated expression. In addition to the single copy of the authentic SOD gene, the genome of H. cutirubrum contains a sequence that is very closely related to but does not code for the previously purified SOD of this organism.

Synthetic promoters are important for temporal and spatial gene expression in transgenic plants. To identify novel microbe-associated molecular pattern (MAMP)-responsive cis-regulatory sequences for synthetic promoter design, a combination of bioinformatics and experimental approaches was employed. One cis-sequence was identified which confers strong MAMP-responsive reporter gene activity with low background activity. The 35-bp-long cis-sequence was identified in the promoter of the Arabidopsis thaliana DJ1E gene, a homologue of the human oncogene DJ1. In this study, this cis-sequence is shown to be a tripartite cis-regulatory module (CRM). A synthetic promoter with four copies of the CRM linked to a minimal promoter increases MAMP-responsive reporter gene expression compared to the wild-type DJ1E promoter. The CRM consists of two WT-boxes (GGACTTTT and GGACTTTG) and a variant of the GCC-box (GCCACC), all required for MAMP and salicylic acid (SA) responsivity. Yeast one-hybrid screenings using a transcription factor (TF)-only prey library identified two AP2/ERFs, ORA59 and ERF10, interacting antagonistically with the CRM. ORA59 activates reporter gene activity and requires the consensus core sequence GCCNCC for gene expression activation. ERF10 down-regulates MAMP-responsive gene expression. No TFs interacting with the WT-boxes GGACTTTT and GGACTTTG were selected in yeast one-hybrid screenings with the TF-only prey library. In transgenic Arabidopsis, the synthetic promoter confers strong and specific reporter gene activity in response to biotrophs and necrotrophs as well as SA.

We have previously isolated a partial cDNA clone encoding a heat shock protein which has been termed hsp 108 (Zarucki-Schulz, T., Kulomaa, M. S., Headon, D. R., Weigel, N. L., Baez, M., Edwards, D. O., McGuire, W. L., Schrader, W. T., and O'Malley, B. W. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 6358-6362; Sargan, D. R., Tsai, M.-J., and O'Malley, B. W. (1985) Biochemistry 25, 6252-6258). Here we examine the expression of the hsp 108 gene in steroid-stimulated chick oviducts. After 16 h of secondary stimulation with estrogen or progesterone, a 20-50-fold increase in hsp 108 mRNA is detected above unstimulated levels. RNA quantitation by Rot analysis shows that in these oviducts there are 75 molecules of hsp 108 mRNA/oviduct cell. Nuclear "run-off" assays indicate only a 2-4-fold increase in the rate of transcription of the gene in response to either sex steroid, suggesting that the gene is regulated both at the transcriptional level and by mRNA stabilization. On hormone withdrawal, the concentration

of hsp 108 mRNA in the oviduct falls to unstimulated control levels within 4 days. Chronic stimulation of the chicks with estrogen (or high acute doses of estrogen) attenuates specifically the inductive response of the hsp 108 gene, but not of ovalbumin, under these conditions. This is not due to a significant reduction of the transcription rate of the gene. We have previously shown that hsp 108 is expressed constitutively in many tissues of the chick (Sargan, D. R., Tsai, M.-J., and O'Malley, B. W. (1985) Biochemistry 25, 6252-6258). In tissues which are not responsive to hormones, no short-term effects of hormone administration on the gene were observed. In the spleen there is a reproducible slow activation of the gene, but the kinetics of this response suggest that it is not a primary response to the hormone. Thus, this hsp 108 gene codes for an interesting new eucaryotic heat shock protein which is regulated also by steroid hormones in a tissue-specific manner

Thiostrepton induced the expression of four proteins (17, 19, 30, and 56 kilodaltons) of unknown function in Streptomyces lividans. The chromosomal gene which encoded the 19-kilodalton protein (tipA) was cloned and sequenced. Transcription of the tipA promoter was induced at least 200-fold by thiostrepton. The tipA 200-fold by thiostrepton. The tipA transcriptional start site (located by S1 mapping and primer extension experiments) was preceded by a 45-base-pair imperfect inverted-repeat sequence which included the -10 and -35 regions of the promoter. Under noninducing conditions in vivo, this might form a cruciform structure which is not recognized by RNA polymerase. A 143-base-pair fragment including this region was cloned into a promoter probe vector, pIJ486. In this plasmid, pAK114, the thiostrepton-inducible tipA promoter controlled the expression of a kanamycin resistance gene encoding an aminoglycoside phosphotransferase. As little as 1 ng of thiostrepton spotted on a lawn of S. lividans(pAK114) induced kanamycin-resistant growth. Other thiostreptonlike antibiotics also induced tipA, but structurally unrelated antibiotics which inhibit translation had no effect. In S. lividans, the promoter could be induced by thiostrepton during either growth or stationary phase. The tipA promoter should be a valuable tool for expression studies in streptomycetes.

The engineering of conditional alleles has evolved from simple floxing of regions of genes to more elaborate methods. Previously, we developed Conditional by Inversion (COIN), an allele design that utilizes an exon-splitting intron and an invertible genetrap-like module (COIN module) to create null alleles upon Cre-mediated inversion. Here we build upon COINs by generating a new Multifunctional Allele (MFA), that utilizes a single gene-targeting step and three site-specific recombination systems, to generate four allelic states: 1. The initial MFA (generated upon targeting) functions as a null with reporter (plus drug selection cassette) allele, wherein the gene of interest is inactivated by both inversion of a critical region of its coding sequence and simultaneous insertion of a reporter gene. MFAs can also be used as 'reverse-conditional' alleles as they are functionally wild type when they are converted to COIN alleles. 2. Null with reporter (minus drug selection cassette), wherein the selection cassette, the inverted critical region, and the COIN module are removed. 3. COIN-based conditional-null via removal of the selection cassette and reporter and simultaneous re-inversion of the critical region of the target. 4. Inverted COIN allele, wherein the COIN allele in turn is reconverted to a null allele by taking advantage of the COIN module's gene trap while simultaneously deleting the critical region.

Genes essential for bacterial growth are of particular scientific interest. Many putative essential genes have been identified or predicted in several species, however, little is known about gene expression requirement stringency, which may be an important aspect of bacterial physiology and likely a determining factor in drug target development.

Mutations in the yeast gene CYH2 can lead to resistance to cycloheximide, an inhibitor of eukaryotic protein synthesis. The gene product of CYH2 is ribosomal protein L29, a component of the 60S ribosomal subunit. We have cloned the wild-type and resistance alleles of CYH2 and determined their nucleotide sequence. Transcription of CYH2 appears to initiate and terminate at multiple sites, as judged by S1 nuclease analysis. The gene is transcribed into an RNA molecule of about 1082 nucleotides, containing an intervening sequence of 510 nucleotides. The splice junction of the intron resides within a codon near the 5' end of the gene. In confirmation of peptide analysis by Stocklein et al. (1) we find

that resistance to cycloheximide is due to a transversion mutation resulting in the replacement of a glutamine by glutamic acid in position 37 of L29.

The CYC7 gene of Saccharomyces cerevisiae encodes the minor species, iso-2, of the cytochrome c protein. Its expression is governed by two regulatory sequences upstream from the gene: a positive site which stimulates transcription 240 base pairs 5' from the protein-coding sequence (-240) and a negative site which inhibits transcription at -300. In this study, the nature of the positive site and its relationship to the negative site has been investigated. Expression of the CYC7 gene is weakly inducible by oxygen. This effect was greatly enhanced by the semidominant CYP1-16 mutation in the trans-acting gene CYP1. The weak oxygen regulation in wild-type cells and the enhanced induction in CYP1-16 mutants were found to be mediated through the positive site. A mutational analysis of this site implicated at least part of a tandem, direct repeat of 9 base pairs as essential for the functioning of this site. The relationship between the positive and negative sites was investigated by comparing the expression of the intact gene with that of derivatives lacking either one or the other site. The expression of the gene containing only the negative site was actually stimulated anaerobically, while the gene containing the positive site alone, although having higher expression aerobically than anaerobically, had higher anaerobic expression than did the intact gene. Thus, it appeared that the combination of the positive and negative sites suppressed anaerobic expression. A model which attempts to explain these properties of the two sites and account for the regulation of the expression of the intact gene is presented.

A human gene encoding an interferon-induced 15-kDa protein has been isolated from a genomic library. The gene appears to be single-copy and is composed of two exons, the first of which contains the ATG translation initiation codon. In vitro nuclear run-on assays showed that the transcription rate of the gene is stimulated after interferon treatment. To analyze transcriptional regulatory sequences, we constructed recombinant plasmids for use in transient transfection assays of HeLa cells. Constructs containing 115 nucleotides 5' to the transcription initiation site were found to be fully inducible by interferon. Assays of deletion mutants identified a critical element for interferon induction located between -115 and -96, just upstream of the "CCAAT box." Moreover, a DNA fragment including this region can confer interferon inducibility on a heterologous promoter (thymidine kinase) when cloned in either orientation upstream of the gene or downstream of the gene. These are properties characteristic of an enhancer element that is active only after treatment with interferon. This regulatory sequence may be shared by a group of interferon-induced genes, since a very similar sequence is present within the functional region near the RNA start site of another interferon-induced gene.

Characterizing the relationship between genetic, epigenetic (e.g., deoxyribonucleic acid [DNA] methylation), and transcript variation could provide insights into mechanisms regulating hemostasis and potentially identify new drug targets. Several hemostatic factors are synthesized in the liver, yet high-resolution DNA methylation data from human liver tissue is currently lacking for these genes. Single-nucleotide polymorphisms (SNPs) can influence DNA methylation in *cis* which can affect gene expression. This can be analyzed through allele-specific methylation (ASM) experiments. We performed targeted genomic DNA- and bisulfite-sequencing of 35 hemostatic genes in human liver samples for SNP and DNA methylation analysis, respectively, and integrated the data for ASM determination. ASM-associated SNPs (ASM-SNPs) were tested for association to gene expression in liver using in-house generated ribonucleic acid-sequencing data. We then assessed whether ASM-SNPs associated with gene expression, plasma proteins, or other traits relevant for hemostasis using publicly available data. We identified 112 candidate ASM-SNPs. Of these, 68% were associated with expression of their respective genes in human liver or in other human tissues and 54% were associated with the respective plasma protein levels, activity, or other relevant hemostatic genome-wide association study traits such as venous thromboembolism, coronary artery disease, stroke, and warfarin dose maintenance. Our study provides the first detailed map of the DNA methylation landscape and ASM analysis of hemostatic genes in human liver tissue, and suggests that methylation regulated by genetic variants in *cis* may provide a mechanistic link between noncoding SNPs and variation observed in circulating hemostatic proteins, prothrombotic diseases, and drug response.

A second, expressed thrombospondin (TSP) gene, Thbs2, has been identified in the mouse. The exon/intron organization of Thbs2 is highly conserved in comparison with Thbs1 in that exon size and the pattern of interruption of the reading frame by introns are preserved, but there is a marked divergence in coding sequence, primarily in the first 7 exons. On the other hand, the DNA and translated amino acid sequences of exons coding for the type I, II, and III repeats in the two TSPs are far better conserved. Thbs2 is located on chromosome 17, band A3, whereas Thbs1 was found on chromosome 2, band F. In marked contrast to Thbs1, the Thbs2 gene is not induced by serum in NIH 3T3 cells; promoter sequences in the two genes are also very different. It is therefore likely that the two TSPs perform related but distinct functions.

We have investigated the basis for liver-specific and sex-linked expression of hepatitis B surface antigen (HBsAg) gene in transgenic mice by monitoring the level of liver HBsAg mRNA and serum HBsAg at different stages of development and in response to sex-hormone regulation. Transcription of the HBsAg gene starts at day 15 of development, together with that of the albumin gene, and reaches a comparable level at birth. HBsAg mRNA level and HBsAg production are parallel in males and females during prenatal development and until the first month of life, but HBsAg gene expression increases 5-10 times in males at puberty. After castration, the level of expression decreases dramatically in both males and females and is subsequently increased by injection of testosterone or estradiol. Glucocorticoids also regulated positively expression of the HBsAg gene. Our results suggest that sex hormones play a role in hepatitis B virus gene expression during natural infection and could explain the difference in incidence of chronic carriers between men and women.

Glucocorticoid regulates various physiological processes via the activation and repression of gene expression. The anti-inflammatory effects and the adverse effects are believed to be dependent on the repression and the activation of genes, respectively. Reporter gene assay is a useful technique to separately evaluate these two functions and has been used for in vitro screening of novel ligands for the glucocorticoid receptor (GR). We report here the application of a reporter gene assay for the in vivo determination of the GR-mediated gene activation. A reporter plasmid containing glucocorticoid response elements was introduced to abdominal mouse skin using a gene gun. Administration of prednisolone induced the expression of the reporter gene, only when the GR expression plasmid was co-transfected with the reporter plasmid. Endogenous levels of corticosterone appeared to be negligible in this protocol. The dose response for this induction was comparable to those for the decreases in thymus weight and serum corticosterone. These results suggest that gene gun-mediated skin transfection enables the in vivo reporter gene assay and that this technique can be used to predict the potency of ligands for the GR-mediated gene activation.

Domoic acid (DA) is a neuroexcitatory amino acid that is naturally produced by some marine diatom species of the genus Pseudo-nitzschia. Ingestion of DA-contaminated seafood by humans results in a severe neurotoxic disease known as amnesic shellfish poisoning (ASP). Clinical signs of ASP include seizures and neuronal damage from activation of ionotropic glutamate receptors. However, the impacts of DA exposure at levels below those known to induce outward signs of neurobehavioral exicitotoxicity have not been well characterized. To further understand the mechanisms of neurotoxic injury associated with DA exposure, we examined the transcriptome of whole brains from zebrafish (Danio rerio) receiving intracoelomic (IC) injection of DA at both symptomatic and asymptomatic doses. A majority of zebrafish exposed to high-dose DA (1.2 microg DA/g) exhibited clinical signs of neuroexcitotoxicity (EC(50) of 0.86 microg DA/g) within 5-20 min of IC injection. All zebrafish receiving low-dose DA (0.47 microg DA/g) or vehicle only maintained normal behavior. Microarray analysis of symptomatic and asymptomatic exposures collectively yielded 306 differentially expressed genes (1.5-fold, p </= 0.05) predominately represented by signal transduction, ion transport, and transcription factor functional categories. Transcriptional profiles were suggestive of neuronal apoptosis following an overwhelming of protective adaptive pathways. Further, potential molecular biomarkers of neuropathic injury, including the zebrafish homolog of human NDRG4, were identified and may be relevant to DA exposure levels below that causing neurobehavioral injury. In general, DA-modulated gene expression

was consistent with other model species thereby validating zebrafish as an appropriate vertebrate model to study mechanisms of DA neurotoxicity. These data provide a basis for identifying pathways of DA-induced injury as well as biomarkers of asymptomatic and symptomatic DA exposure levels.

Many questions about the biological activity and availability of small molecules remain inaccessible to investigators who could most benefit from their answers. To narrow the gap between chemoinformatics and biology, we have developed a suite of ligand annotation, purchasability, target, and biology association tools, incorporated into ZINC and meant for investigators who are not computer specialists. The new version contains over 120 million purchasable "drug-like" compounds--effectively all organic molecules that are for sale--a quarter of which are available for immediate delivery. ZINC connects purchasable compounds to high-value ones such as metabolites, drugs, natural products, and annotated compounds from the literature. Compounds may be accessed by the genes for which they are annotated as well as the major and minor target classes to which those genes belong. It offers new analysis tools that are easy for nonspecialists yet with few limitations for experts. ZINC retains its original 3D roots--all molecules are available in biologically relevant, ready-to-dock formats.

The phytohormone, abscisic acid (ABA), plays a variety of roles during seed development and in the plant's response to environmental stresses. To study the molecular action of ABA, we have isolated a single copy ABA-induced gene, HVA22, which is mapped to barley chromosome 1. The HVA22 gene can be induced by either ABA or the protein synthesis inhibitor, cycloheximide, and addition of both inducers to barley aleurone layers has a synergistic effect on the expression of this gene. Sequence comparison indicates that the HVA22 gene product is highly homologous to the product of human DP1 gene, which is likely to contribute to colorectal tumorigenesis. The hormonal regulation of HVA22 expression has been studied, and there appear to be at least three elements, two located in the promoter and one in the first intron, which are essential for the high level of ABA induction of HVA22 expression. Among the promoter elements is a homolog of ABA response element, which has been shown to be important in the expression of other ABA-induced genes in plants. We suggest that the barley HVA22 gene product is likely a regulatory protein, and the ABA induction of this gene requires the action of a complex set of hormone response elements.

To study the mechanism of induction of human C-reactive protein (CRP) gene expression, we have utilized an in vitro liver cell system to analyze the cis-acting DNA sequences located within the 5'-flanking region of human CRP gene. Stable transfection of human hepatoma cells, PLC/PRF/5, by a CRP gene construct containing the 1 kilobase pair of upstream sequence of the CRP gene demonstrated that this region contained the inducible element(s) which regulated human CRP gene transcription. Dissection of this region by 5', 3' and internal deletion constructs of upstream region of the CRP gene fused to a reporter gene, chloramphenicol acetyl transferase, indicated the presence of two inducible elements located proximal to the site of initiation of transcription, two constitutive enhancer-like elements located distal to the promoter, and a negative regulatory region located between the two inducible elements. We had previously shown that a protein factor from monocytes or HTLV1-infected T-cells, was responsible for CRP induction in hepatoma cells. We have found this factor to be synonymous with interleukin-6. By stable and transient transfection assays in hepatoma cells, recombinant interleukin-6 alone was sufficient to activate both inducible elements.

beta-Interferon (beta-IFN) gene expression can be induced by poly(I)-poly(C) or virus, but there is considerable variation in the extent of induction between different cell lines. We characterized two poorly inducible human cell lines, HeLa and 143 thymidine kinase negative (143 tk-), to define cellular factors involved in the activation of the beta-IFN gene. We show that the deficiency in beta-IFN induction in these cells can be complemented by fusion to highly inducible mouse cells. We conclude that the human cells are deficient in a trans-acting factor required for B-IFN gene activation. The level of induction of the beta-IFN gene in HeLa and 143 tk- cells can also be increased by priming with IFN before induction. If IFN priming is carried out in the presence of cycloheximide, a approximately 200-

fold increase in induction is observed. We conclude that activation of the beta-IFN gene requires an IFN-inducible factor that is only expressed at low levels in unprimed HeLa and 143 tk- cells.

The ability to regulate the expression of a gene greatly aids the process of uncovering its functions. The fission yeast Schizosaccharomyces pombe has so far lacked a system for rapidly controlling the expression of chromosomal genes, hindering its full potential as a model organism. Although the widely used nmt1 promoter displays a wide dynamic range of activity, it takes > 14-15 h to derepress. The urg1 promoter also shows a large dynamic range and can be induced quickly (< 2 h), but its implementation requires laborious strain construction and it cannot be used to study meiosis. To overcome these limitations, we constructed a tetracycline-regulated system for inducible expression of chromosomal genes in fission yeast, which is easily established and implemented. In this system the promoter of a gene is replaced by simple one-step substitution techniques with a tetracycline-regulated promoter cassette (tetO(7) -TATA(CYC1)) in cells where TetR/TetR'-based transcription activators/repressors are also produced. Using top1 and nse6 as reporter genes, we show that Top1 and Nse6 appear after just 30 min of activating tetO(7) -TATA(CYC1) and plateau after -4-6 h. The amount of synthesised protein is comparable to that produced from the attenuated nmt1 promoter P(nmt8) , which should be closer to wild-type levels for most genes than those generated from excessively strong promoters and can be controlled by changing the concentration of the effector antibiotic. This system also works efficiently during meiosis, thus making it a useful addition to the toolkit of the fission yeast community.

The gene transfer agent (GTA) is a phage-like particle capable of exchanging double-stranded DNA fragments between cells of the photosynthetic bacterium Rhodobacter capsulatus. Here we show that the major capsid protein of GTA, expressed in E. coli, can be assembled into prohead-like structures in the presence of calcium ions in vitro. Transmission electron microscopy (TEM) of uranyl acetate staining material and thin sections of glutaraldehyde-fixed material demonstrates that these associates have spherical structures with diameters in the range of 27-35 nm. The analysis of scanning TEM images revealed particles of mass approximately 4.3 MDa, representing 101+/-11 copies of the monomeric subunit. The establishment of this simple and rapid method to form prohead-like particles permits the GTA system to be used for genome manipulation within the photosynthetic bacterium, for specific targeted drug delivery, and for the construction of biologically based distributed autonomous sensors for environmental monitoring.

Sequence mutations represent a driving force of adaptive evolution in bacterial pathogens. It is especially evident in reductive genome evolution where bacteria underwent lifestyles shifting from a free-living to a strictly intracellular or host-depending life. It resulted in loss-of-function mutations and/or the acquisition of virulence gene clusters. Bacillus anthracis shares a common soil bacterial ancestor with its closely related bacillus species but is the only obligate, causative agent of inhalation anthrax within the genus Bacillus. The anthrax-causing Bacillus anthracis experienced the similar lifestyle changes. We thus hypothesized that the bacterial pathogen would follow a compatible evolution path.

The recombinant baculovirus has been widely used as an efficient tool to mediate gene delivery into mammalian cells but has barely been used in fish cells. In the present study, we constructed a recombinant baculovirus containing the dual-promoter cytomegalovirus (CMV) and white spot syndrome virus (WSSV) immediate-early gene 1 (ie1) (WSSV ie1), followed by a puromycin⁻green fluorescent protein (Puro-GFP, *pf*) or puromycin⁻red fluorescent protein (Puro-RFP, *pr*) cassette, which simultaneously allowed for easy observation, rapid titer determination, drug selection, and exogenous gene expression. This recombinant baculovirus was successfully transduced into fish cells, including *Mylopharyngodon piceus* bladder (MPB), fin (MPF), and kidney (MPK); *Oryzias latipes* spermatogonia (SG3); and *Danio rerio* embryonic fibroblast (ZF4) cells. Stable transgenic cell lines were generated after drug selection, which was further verified by Western blot. A cell monoclonal formation assay proved the stable heredity of transgenic MPB cells. In addition, a recombinant baculovirus containing a *pr* cassette and four transcription factors for induced pluripotent stem cells

(iPSC) was constructed and transduced into ZF4 cells, and these exogenous genes were simultaneously delivered and transcribed efficiently in drug-selected ZF4 cells, proving the practicability of this modified recombinant baculovirus system. We also proved that the WSSV ie1 promoter had robust activity in fish cells in vitro and in vivo. Taken together, this modified recombinant baculovirus can be a favorable transgenic tool to obtain transient or stable transgenic fish cells.

Multiphoton-targeted photochemistry was used to selectively inactivate the expression of genes in vertebrate cells. A membrane permeable DNA-associating vital dye, ethidium bromide monoacetate (visible wavelength single photon absorption peak at 530 nm) was used to photosensitize chromosomes in dividing cells. A 100-ps infrared laser beam operating at 1.06 microns was focused onto a selected region of a mitotic chromosome corresponding to the sites of the nucleolar (ribosomal) genes. Individual cells followed through mitosis demonstrated a reduction in the number of nucleoli formed in daughter cells that corresponded to the number of nucleolar genes sites irradiated. These results demonstrate the ability to selectively manipulate genes by using the focal point specificity characteristic of multiphoton microscopy. This technique should have wide biotechnology applications both in vitro and in vivo.

Using the relative expression levels of two SNP alleles of a gene in the same sample is an effective approach for identifying cis-acting regulatory SNPs (rSNPs). In the current study, we established a process for systematic screening for cis-acting rSNPs using experimental detection of AI as an initial approach. We selected 160 expressed candidate genes that are involved in cancer and anticancer drug resistance for analysis of AI in a panel of cell lines that represent different types of cancers and have been well characterized for their response patterns against anticancer drugs. Of these genes, 60 contained heterozygous SNPs in their coding regions, and 41 of the genes displayed imbalanced expression of the two cSNP alleles. Genes that displayed AI were subjected to bioinformatics-assisted identification of rSNPs that alter the strength of transcription factor binding. rSNPs in 15 genes were subjected to electrophoretic mobility shift assay, and in eight of these genes (APC, BCL2, CCND2, MLH1, PARP1, SLIT2, YES1, XRCC1) we identified differential protein binding from a nuclear extract between the SNP alleles. The screening process allowed us to zoom in from 160 candidate genes to eight genes that may contain functional rSNPs in their promoter regions

Nearly 20% of yeast genes are required for viability, hindering genetic analysis with knockouts. We created promoter-shutoff strains for over two-thirds of all essential yeast genes and subjected them to morphological analysis, size profiling, drug sensitivity screening, and microarray expression profiling. We then used this compendium of data to ask which phenotypic features characterized different functional classes and used these to infer potential functions for uncharacterized genes. We identified genes involved in ribosome biogenesis (HAS1, URB1, and URB2), protein secretion (SEC39), mitochondrial import (MIM1), and tRNA charging (GSN1). In addition, apparent negative feedback transcriptional regulation of both ribosome biogenesis and the proteasome was observed. We furthermore show that these strains are compatible with automated genetic analysis. This study underscores the importance of analyzing mutant phenotypes and provides a resource to complement the yeast knockout collection.

The NAD(P)H:menadione oxidoreductase gene (Nmo-1) codes for a quinone reductase (also called DT diaphorase; EC 1.6.99.2) believed to play a central role in protection against oxidative stress. We have studied mice with a radiation-induced chromosomal deletion involving the albino locus (c) on chromosome 7 and found that Nmo-1 mRNA levels and the rate of Nmo-1 gene transcription are markedly increased (greater than 100-fold and greater than 12-fold, respectively) in the untreated c14CoS/c14CoS deletion homozygote, compared with the untreated Cch/Cch wild-type and the Cch/C14CoS heterozygote. These data suggest that a gene located on chromosome 7 encodes a trans-acting regulatory factor that might be a negative effector of the Nmo-1 gene, which we show here is located on chromosome 8 approximately 1.4 centimorgans (about 1000 kilobase pairs) from the Es-2 gene. Conversely, there are no detectable basal levels of cytochrome P1450 (Cyp1a1 gene) or

cytochrome P3450 (Cyp1a2 gene) mRNA, indicating that the regulation of basal expression of the Cyp1a1 and Cyp1a2 genes is distinct from that of the Nmo-1 gene. Moreover, the Cyp1a1 and Cyp1a2 genes and the Nmo-1 gene are induced by tetrachlorodibenzo-p-dioxin in the cch/cch, cch/c14CoS, and c14CoS/c14CoS mice. The mechanism of tetrachlorodibenzo-p-dioxin inducibility of the Cyp1a1, Cyp1a2, and Nmo-1 genes is, therefore, independent of the mechanism of Nmo-1 gene activation in untreated c14CoS/c14CoS mice.

The methionine component of glyA gene regulation in Escherichia coli K-12 was investigated. The results indicate that the glyA gene is positively controlled by the metR gene product. Activation of glyA by the MetR protein requires homocysteine, an intermediate in methionine biosynthesis. The positive-acting metR regulatory system functions independently of a regulatory system shown previously to control glyA gene expression.

Cancer mutation databases are expected to play central roles in personalized medicine by providing targets for drug development and biomarkers to tailor treatments to each patient. The accuracy of reported mutations is a critical issue that is commonly overlooked, which leads to mutation databases that include a sizable number of spurious mutations, either sequencing errors or passenger mutations. Here we report an analysis of the latest version of the TP53 mutation database, including 34,453 mutations. By using several data-driven methods on multiple independent quality criteria, we obtained a quality score for each report contributing to the database. This score can now be used to filter for high-confidence mutations and reports within the database. Sequencing the entire TP53 gene from various types of cancer using next-generation sequencing with ultradeep coverage validated our approach for curation. In summary, 9.7% of all collected studies, mostly comprising numerous tumors with multiple infrequent TP53 mutations, should be excluded when analyzing TP53 mutations. Thus, by combining statistical and experimental analyses, we provide a curated mutation database for TP53 mutations and a framework for mutation database analysis.

Candida albicans, the single most frequently isolated human fungal pathogen, was thought to be asexual until the recent discovery of the mating-type-like locus (MTL). Homozygous MTL strains were constructed and shown to mate. Furthermore, it has been demonstrated that opaque-phase cells are more efficient in mating than white-phase cells. The similarity of the genes involved in the mating pathway in Saccharomyces cerevisiae and C. albicans includes at least one gene (KEX2) that is involved in the processing of the alpha mating pheromone in the two yeasts. Taking into account this similarity, we searched the C. albicans genome for sequences that would encode the alpha pheromone gene. Here we report the isolation and characterization of the gene MFalpha1, which codes for the precursor of the alpha mating pheromone in C. albicans. Two active alpha-peptides, 13 and 14 amino acids long, would be generated after the precursor molecule is processed in C. albicans. To examine the role of this gene in mating, we constructed an mfalpha1 null mutant of C. albicans. The mfalpha1 null mutant fails to mate as MTLalpha, while MTLa mfalpha1 cells are still mating competent. Experiments performed with the synthetic alpha-peptides show that they are capable of inducing growth arrest, as demonstrated by halo tests, and also induce shmooing in MTLa cells of C. albicans. These peptides are also able to complement the mating defect of an MTLalpha kex2 mutant strain when added exogenously, thereby confirming their roles as alpha mating pheromones.

Nucleotide sequencing of a human cosmid clone shows that the exon-intron structures of a glutathione S-transferase multigene family are conserved between man and rat, that the human gene family is clustered and that gene conversion events have occurred within the cluster. In addition, between man and rat, there is a high degree of nucleotide sequence identity not only in exons but also in some introns. These conserved sequences are coincident with homologous sequences subject to gene conversion in both species, and hence the utilization of gene conversion by this gene family has itself been conserved. By using transient-expression assay the conserved/converted regions are shown to be capable of modulating transcriptional activity. The data suggest that DNA repair by gene conversion may be a

chemical immunity mechanism. which could result in acquired resistance to toxins and, in particular, drug resistance due to glutathione S-transferase in tumours.

Negative selectable markers are useful tools for forward-genetic screens aimed at identifying trans-acting factors that are required for expression of specific genes. Transgenic lines harbouring the marker fused to a gene element, such as a promoter, may be mutagenized to isolate loss-of-function mutants able to survive under selection. Such a strategy allows the molecular dissection of factors that are essential for expression of the gene. Expression of individual chloroplast genes in plants and algae typically requires one or more nuclear-encoded factors that act at the post-transcriptional level, often through interaction with the 5' UTR of the mRNA. To study such nuclear control further, we have developed the Escherichia coli cytosine deaminase gene codA as a conditional negative selectable marker for use in the model green alga Chlamydomonas reinhardtii. We show that a codon-optimized variant of codA with three amino acid substitutions confers sensitivity to 5-fluorocytosine (5-FC) when expressed in the chloroplast under the control of endogenous promoter/5' UTR elements from the photosynthetic genes psaA or petA. UV mutagenesis of the psaA transgenic line allowed recovery of 5-FC-resistant, photosynthetically deficient lines harbouring mutations in the nuclear gene for the factor TAA1 that is required for psaA translation. Similarly, the petA line was used to isolate mutants of the petA mRNA stability factor MCA1 and the translation factor TCA1. The codA marker may be used to identify critical residues in known nuclear factors and to aid the discovery of additional factors required for expression of chloroplast genes.

We constructed an assay system of a luciferase reporter with p16/lNK4a gene transcriptional regulatory domain to identify p16-inducing substances, and found toyocamycin to induce gene expression from the screening of culture fluids of Streptomyces. Toyocamycin is a nucleoside analog, and it increased the p16 mRNA level in human normal fibroblasts or synovial cells as assessed by Northern blot hybridization or real time RT-PCR. It also induced cellular senescence in normal human fibroblasts. The transcriptional regulatory regions of human p16 gene that were responsible for the induction were analyzed using deletion mutants of the transcriptional regulatory region of p16 linked to the luciferase gene. The DNA fragment -111 to +1 bp from the cap site was sufficient to drive toyocamycin-activated transcription of p16/luciferase reporter. Nucleotide sequences within this domain contained the Sp1- and Ets-binding sequences. Mutations were introduced into these sequences, and the Sp1 sequence was found to be critical for the induction, and this notion was confirmed from gel-mobility shift assay.

Increased expression of major histocompatibility complex (MHC) class-I genes and aberrant expression of MHC class-II genes in thyroid epithelial cells (TECs) are associated with autoimmune thyroid diseases. Previous studies have shown that methimazole (MMI) reduces MHC class-I expression and inhibits interferon-gamma (IFN-gamma or IFNG as listed in the MGI Database)-induced expression of the MHC class-II genes in TECs. The action of MMI on the MHC class-I genes is transcriptional, but its mechanism has not been investigated previously. In the present study, we show that in Fisher rat thyroid cell line 5 cells, the ability of MMI and its novel derivative phenylmethimazole (C10) to decrease MHC class-I promoter activity is similar to TSH/cAMP suppression of MHC class-I and TSH receptor genes, and involves a 39 bp silencer containing a cAMP response element (CRE)-like site. Furthermore, we show that C10 decreases MHC class-I gene expression to a greater extent than MMI and at 10- to 50-fold lower concentrations. C10 also reduces the IFN-gamma-induced increase in the expression of MHC class-I and MHC class-II genes more effectively than MMI. Finally, we show that in comparison to MMI, C10 is a better inhibitor of specific protein-DNA complexes that are formed with a CRE-like element on the MHC class-II promoter. These data support the conclusion that the immunosuppressive mechanism by which MMI and C10 inhibit MHC gene expression mimics 'normal' hormonal suppression by TSH/cAMP.

The ability of hepatitis B virus (HBV) to stimulate the expression of a cellular gene was investigated by using a transient-expression system. A plasmid in which the expression of the bacterial chloramphenicol acetyltransferase (cat) gene had been placed under the control of the DNA sequences

that regulate the expression of the human beta-interferon gene was constructed. In Vero cells, cotransfection of the 2.7-kilobase BglII DNA fragment of HBV together with the test plasmid containing the cat gene resulted in stimulation of the expression of the cat gene. This HBV DNA fragment was specific in its trans-activation; no significant stimulation of CAT activity was observed in constructs when the promoter and enhancer elements were derived from the murine sarcoma viral long terminal repeat, Rous sarcoma virus, BK virus, or simian virus 40. Results of subcloning of the HBV DNA fragment indicate that the trans-activating function resides in a 944-base-pair EcoRV-BglII DNA fragment of the HBV genome that contains the X structural gene and its promoter element. Removal of the promoter from the X structural gene resulted in loss of the trans-activating function. A frameshift mutation within the X gene region also eliminated the trans-activating activity. These results suggest that the X antigen could play a role in HBV infections by activating the expression of cellular genes.

For the transformation of the yeast species Kluyveromyces fragilis, we have constructed a vector containing a bacterial kanamycin resistance (Kmr) gene, the TRP1 gene of Saccharomyces cerevisiae, and an autonomously replicating sequence of Kluyveromyces lactis called KARS2 . By utilizing the method based on treatment by alkali cations and with the Kmr gene as the selective marker, a wild-type strain of K. fragilis was transformed to resistance against the antibiotic G418 . In the transformed cell the plasmid replicates autonomously. The same plasmid could also be used to transform S. cerevisiae trp1 mutant to Trp+. Thus, KARS2 of K. lactis enables the vector to replicate in K. fragilis, K. lactis, and S. cerevisiae, whereas ARS1 of S. cerevisiae allows autonomous replication only in S. cerevisiae.

The mof4-1 (maintenance of frame) allele in the yeast Saccharomyces cerevisiae was isolated as a chromosomal mutation that increased the efficiency of -1 ribosomal frameshifting at the L-A virus frameshift site and caused loss of M1, the satellite virus of L-A. Here, we demonstrate that strains harboring the mof4-1 allele inactivated the nonsense-mediated mRNA decay pathway. The MOF4 gene was shown to be allelic to UPF1, a gene whose product is involved in the nonsense-mediated mRNA decay pathway. Although cells harboring the mof4-1 allele of the UPF1 gene lose the M1 virus, mutations in other UPF genes involved in nonsense-mediated mRNA decay maintain the M1 virus. The mof4-1 strain is more sensitive to the aminoglycoside antibiotic paromomycin than a upf1 delta strain, and frameshifting efficiency increases in a mof4-1 strain grown in the presence of this drug. Further, the ifs1 and ifs2 alleles previously identified as mutations that enhance frameshifting were shown to be allelic to the UPF2 and UPF1 genes, respectively, and both ifs strains maintained M1. These results indicate that mof4-1 is a unique allele of the UPF1 gene and that the gene product of the mof4-1 allele affects both -1 ribosomal frameshifting and mRNA turnover.

Identical genes in the same cellular environment are sometimes expressed differently. In some cases, including the immunoglobulin heavy chain (IgH) locus, this type of differential gene expression has been related to the absence of a transcriptional enhancer. To gain additional information on the role of the IgH enhancer, we examined expression driven by enhancers that were merely weakened, rather than fully deleted, using both mutations and insulators to impair enhancer activity. For this purpose we used a LoxP/Cre system to place a reporter gene at the same genomic site of a stable cell line. Whereas expression of the reporter gene was uniformly high in the presence of the normal, uninsulated enhancer and undetectable in its absence, weakened enhancers yielded variegated expression of the reporter gene; i.e., the average level of expression of the same gene differed in different clones, and expression varied significantly among cells within individual clones. These results indicate that the weakened enhancer allows the reporter gene to exist in at least two states. Subtle aspects of the variegation suggest that the IgH enhancer decreases the average duration (half-life) of the silent state. This analysis has also tested the conventional wisdom that enhancer activity is independent of distance and orientation. Thus, our analysis of mutant (truncated) forms of the IgH enhancer revealed that the 250 bp core enhancer was active in its normal position, approximately 1.4 kb 3' of the promoter, but inactive approximately 6 kb 3', indicating that the activity of the core enhancer was distance-dependent. A longer segment--the core enhancer plus approximately 1 kb of 3' flanking material, including the 3' matrix attachment region--

was active, and the activity of this longer segment was orientation-dependent. Our data suggest that this 3' flank includes binding sites for at least two activators.

It has been known that enzyme activity associated with the yeast LEU1 and LEU2 gene product (beta-isopropylmalate dehydrogenase) drops sharply when yeast is grown in the presence of leucine. RNA blot hybridizations with LEU2-specific probes establish that this is accompanied by a 5-fold repression in LEU2 mRNA levels. A similar repression was noted recently for LEU1 mRNA levels (Hsu, Y.-P., and Schimmel, P. (1984) J. Biol. Chem. 259, 3714-3719). Nuclease mapping of the 5'-end of the LEU2 mRNA shows a major start at approximately 16 nucleotides upstream of the AUG initiation codon. This initiation site in the gene is retained in an extensive LEU2 5'-noncoding region deletion which still expresses the LEU2 gene product (Erhart, E., and Hollenberg, C. P. (1983) J. Bacteriol. 156, 625-635). The primary structure of the LEU2 gene product was established from the nucleotide sequence of the gene-coding region and from fitting amino acid sequences of scattered internal peptides to the nucleotide sequence. The 364-amino acid protein has a 13-amino acid stretch which is highly homologous to the partially sequenced yeast LEU1 gene product (isopropylmalate isomerase). The homology occurs about 290 amino acids from the respective NH2 termini of the two proteins. The homology may represent residues which interact with beta-isopropylmalate, a common ligand for the enzymes.

In recent years, the maturation of microarray technology has allowed the genome-wide analysis of gene expression patterns to identify tissue-specific and ubiquitously expressed ('housekeeping') genes. We have performed a functional and topological analysis of housekeeping and tissue-specific networks to identify universally necessary biological processes, and those unique to or characteristic of particular tissues.

The DNA sequence of an aminoglycoside phosphotransferase gene (aph) from Streptomyces fradiae ATCC 10745 (a neomycin producer) was determined. The gene was localized by in vitro subcloning and insertional inactivation. Molecular weight, amino acid composition, and amino-terminal analysis of the purified aph gene product confirmed the accuracy and position of the aph gene sequence. Pairwise comparisons of S. fradiae aph with the aph genes encoded by bacterial transposons Tn5 and Tn903 showed significant nucleotide and amino acid homologies which indicated a common evolutionary origin for these antibiotic-resistance genes.

A variety of genes are involved in determining the level of organic solvent tolerance of Escherichia coli K-12. Gene ostA is one of the genes contributing to the level of organic solvent tolerance. This gene was cloned from an n-hexane-tolerant strain of E. coli, JA300. A JA300-based n-hexane-sensitive strain, OST4251, was converted to the n-hexane-tolerant phenotype by transformation with DNA containing the ostA gene derived from JA300. Thus, the cloned ostA gene complemented the n-hexane-sensitive phenotype

Synthetic lethality is a state when simultaneous loss of two genes is lethal to a cancer cell, while the loss of the individual genes is not. We developed an R package DiscoverSL to predict and visualize synthetic lethality in cancers using multi-omic cancer data. Mutation, copy number alteration and gene expression data from The Cancer Genome Atlas project were combined to develop a multi-parametric Random Forest classifier. The effects of selectively targeting the predicted synthetic lethal genes is tested in silico using shRNA and drug screening data from cancer cell line databases. The clinical outcome in patients with mutation in primary gene and over/under-expression in the synthetic lethal gene is evaluated using Kaplan-Meier analysis. The method helps to identify new therapeutic approaches by exploiting the concept of synthetic lethality.

Molecular responses to genotoxic stress are complex and are mediated by a variety of regulatory pathways. One key element in cellular response is the stress gene transcription factor p53, which can regulate nearly 100 genes that have already been identified. Although p53 plays a central role in the cellular response to DNA-damaging agents such as ionizing radiation (IR), other pathways can also

have important roles. One example is the transcriptional responses associated with IR-induced apoptosis, where induction of some genes is limited to p53 wild-type (wt) cells that also have the ability to undergo rapid apoptosis after irradiation. In contrast, other genes are triggered after IR in lines undergoing rapid apoptosis regardless of p53 status. From this and other examples, it is apparent that the pattern of stress gene expression is cell type specific in both primary and transformed lines. The premise will be developed that such differences in stress gene responsiveness can be employed as molecular markers using a combination of informatics and functional genomics approaches. An example is given using the panel of lines of the NCI anticancer drug screen where both the p53 status and sensitivity to a large collection of cytotoxic agents have been determined. The utility of cDNA microarray hybridization to measure IR-stress gene responses has recently been demonstrated and a large number of additional IR-stress genes have been identified. The responses of some of these genes to IR and other DNA-damaging agents varied widely in cell lines from different tissues of origin and different genetic backgrounds, highlighting the importance of cellular context to genotoxic stress responses; this also highlights the need for informatics approaches to discover and prioritize hypotheses regarding the importance of particular cellular factors. The aim of this review is to demonstrate the utility of combining an informatics approach with functional genomics in the study of stress responses.

The nucleotide sequence of an 8.2 kb BamHI fragment containing the entire chicken histone H2AF gene has been determined. Unlike the majority of histone genes, the coding region is interrupted by four intervening sequences. While sequencing the 8.2 kb BamHI fragment it was found that the promoter and first exon of an unidentified non-histone gene lies immediately downstream of the H2AF gene. Studies of H2AF gene transcription show that, unlike the major core and H1 histone genes, it is not coupled to DNA synthesis.

Analysis of the nucleotide and deduced amino acid sequences of the biologically active mouse mdr1 cDNA clone indicates that the protein is formed by two highly homologous halves, each containing six putative transmembrane domains and a nucleotide-binding site. The duplicated unit shows high sequence homology to the proposed energy-coupling subunit of bacterial periplasmic transport proteins. We have cloned and characterized the mouse mdr1 gene and have analyzed the genomic organization of the two homologous halves forming the mdr1 protein. The gene spans 68 kilobases, is split into 28 exons, and the two homologous halves are encoded by 14 and 13 exons. The transcriptional initiation site of the gene has been mapped and putative TATA and consensus CAAT sequences have been found at positions -27 and -83, respectively. Discrete structural domains of the mdr1 protein are encoded by separate exons: Ten of the 12 putative transmembrane domains are encoded by individual exons and the two nucleotide-binding sites are each encoded by three exons. The exon/intron organization of the gene is conserved in the two highly homologous regions encoding the nucleotide-binding sites. The conservation of certain pairs of introns, together with the high degree of sequence homology, indicate that the mouse mdr1 gene originated from the duplication of an intron-containing ancestral gene.

Programmed cell death (apoptosis) is a normal process by which cells are eliminated during normal embryonic development and in adult life. Disruption of this normal process resulting in illegitimate cell survival can cause developmental abnormalities and facilitate cancer development. Normal cells require certain viability factors and undergo programmed cell death when these factors are withdrawn. The viability factors are required throughout the differentiation process from immature to mature cells. Although many viability factors are also growth factors, viability and growth are separately regulated. Viability factors can have clinical value in decreasing the loss of normal cells including the loss that occurs after irradiation, exposure to other cytotoxic agents or virus infection including AIDS. There is no evidence that occurs after irradiation, exposure to other cytotoxic agents or virus infection including AIDS. There is no evidence that cancer cells are immortal. Programmed cell death can be induced in leukemic cells by removal of viability factors, by cytotoxic therapeutic agents, or by the tumor-suppressor gene wild-type p53. All these forms of induction of programmed cell death in leukemic cells can be suppressed by the same viability factors that suppress programmed cell death in normal cells. A tumor-promoting phorbol ester can also suppress this death program. The induction of programmed cell

death can be enhanced by deregulated expression of the gene c-myc and suppressed by the gene bcl-2. Mutant p53 and bcl-2 suppress the enhancing effect on cell death of deregulated c-myc, and thus allow induction of cell proliferation and inhibition of differentiation which are other functions of deregulated c-myc. The suppression of cell death by mutant p53 and bcl-2 increases the probability of developing cancer. The suppression of programmed cell death in cancer cells by viability factors suggests that decreasing the level of these factors may increase the effectiveness of cytotoxic cancer therapy. Treatments that downregulate the expression or activity of mutant p53 and bcl-2 in cancer cells should also be useful for therapy.

A systematic strategy was used to create a synoptic set of mutations that are distributed throughout the single beta-tubulin gene of Saccharomyces cerevisiae. Clusters of charged amino acids were targeted for mutagenesis and converted to alanine to maximize alterations on the protein's surface and minimize alterations that affect protein folding. Of the 55 mutations we constructed, three confer dominant-lethality, 11 confer recessive-lethality, 10 confer cold-sensitivity, one confers heat-sensitivity, and 27 confer altered resistance to benomyl. Only 11 alleles give no discernible phenotype. In spite of the fact that beta-tubulin is a highly conserved protein, three-fourths of the mutations do not destroy the ability of the protein to support the growth of yeast at 30 degrees C. The lethal substitutions are primarily located in three regions of the protein and presumably identify domains most critical for beta-tubulin function. Interestingly, most of the conditional-lethal alleles produce specific defects in spindle assembly at their restrictive temperature; cytoplasmic microtubules are relatively unaffected. The exceptions are two mutants that contain abnormally long cytoplasmic microtubules. Mutants with specific spindle defects were not observed in our previous collection of beta-tubulin mutants and should be valuable in dissecting spindle function.

The discovery of novel classes of antifungal drugs depends to a certain extent on the identification of new, unexplored targets that are essential for growth of fungal pathogens. Likewise, the broad-spectrum capacity of future antifungals requires the target gene(s) to be conserved among key fungal pathogens. Using a genome comparison (or concordance) tool, we identified 240 conserved genes as candidates for potential antifungal targets in 10 fungal genomes. To facilitate the identification of essential genes in Candida albicans, we developed a repressible C. albicans MET3 (CaMET3) promoter system capable of evaluating gene essentiality on a genome-wide scale. The CaMET3 promoter was found to be highly amenable to controlled gene expression, a prerequisite for use in target-based whole-cell screening. When the expression of the known antifungal target C. albicans ERG1 was reduced via down-regulation of the CaMET3 promoter, the CaERG1 conditional mutant strain became hypersensitive, specifically to its inhibitor, terbinafine. Furthermore, parallel screening against a small compound library using the CaERG1 conditional mutant under normal and repressed conditions uncovered several hypersensitive compound hits. This work therefore demonstrates a streamlined process for proceeding from selection and validation of candidate antifungal targets to screening for specific inhibitors.

We have determined the nucleotide sequence of the ereB gene of plasmid pIP1527 which confers high-level resistance to erythromycin by inactivation in Escherichia coli. The open reading frame of the ereB gene, 1257-bp, was defined by initiation and termination codons and by cloning in vitro. The corresponding protein has a calculated Mr of 48,118 in close agreement with a previous estimation, 51,000, by electrophoresis of minicell extracts in SDS-polyacrylamide gels. The structure of the modified erythromycin was determined by physico-chemical techniques including mass spectrometry, infrared spectrophotometry and 13C nuclear magnetic resonance. The data obtained indicated that like ereA (Ounissi and Courvalin, 1985) ereB encodes an erythromycin esterase. Comparison of the amino acid sequences of the two isozymes did not reveal any statistically significant homology. Analysis of the nucleotide sequence of the ereB gene suggests that this resistance determinant should be exogenous to E. coli.

There are two yeast enolase genes, designated ENO1 and ENO2, which are expressed differentially in vegetative cells grown on glucose and in cells grown on gluconeogenic carbon sources. ENO2 is

induced more than 20-fold in cells grown on glucose, whereas ENO1 expression is similar in cells grown on glucose and in cells grown on gluconeogenic carbon sources. Sequences within the 5' flanking region of ENO2 which are required for glucose-dependent induction were identified by deletion mapping analysis. These studies were carried out by using a fused gene containing the ENO2 5' flanking sequences and the ENO1 coding sequences. This fused gene undergoes glucose-dependent induction and is expressed at the same level as the resident ENO2 gene in cells grown on glucose or gluconeogenic carbon sources. Expression of fused genes containing deletion mutations within the ENO2 5' flanking region was monitored after integration at the ENO1 locus of a strain carrying a deletion of the resident ENO1 coding sequences. This analysis showed that there are two upstream activation sites located immediately upstream and downstream from a position 461 base pairs upstream from the transcriptional initiation site. Either one of these upstream activation sites is sufficient for glucose-dependent induction and normal gene expression in the presence of gluconeogenic carbon sources. Deletion of both regulatory regions results in a complete loss of gene expression. The regulatory regions function normally in both orientations relative to the coding sequences. Mutant fused genes containing small deletions within the regulatory regions were constructed; these genes were expressed normally in gluconeogenic carbon sources but were not induced in the presence of glucose. Based on this analysis, ENO2 contains a cis-acting regulatory region which is required for gene expression and mediates glucose-dependent induction of gene expression.

Btg2 is a primary p53 transcriptional target gene which may function as a coactivator-corepressor and/or an adaptor molecule that modulates the activities of its interacting proteins. We have generated Btg2-null mice to elucidate the in vivo function of Btg2. Btg2-null mice are viable and fertile but exhibit posterior homeotic transformations of the axial vertebrae in a dose-dependent manner. Consistent with its role in vertebral patterning, Btg2 is expressed in the presomitic mesoderm, tail bud, and somites during somitogenesis. We further provide biochemical evidence that Btg2 interacts with bone morphogenetic protein (BMP)-activated Smads and enhances the transcriptional activity of BMP signaling. In view of the genetic evidence that reduced BMP signaling causes posteriorization of the vertebral pattern, we propose that the observed vertebral phenotype in Btg2-null mice is due to attenuated BMP signaling.

We have characterized two new transgenic Xenopus lines enabling transgene expression using the Tet-On inducible system. An inducer line expresses the doxycycline- (Dox-) activated transcription factor rtTA under control of the ubiquitous promoter CMV. A responder line enables Dox-inducible expression of a dominant positive thyroid hormone receptor via a tetracycline responsive transgenic promoter (TRE). Dox-induced expression of transgenic GFP mRNA was detectable after 3 hr and increased up to 10- to 50-fold by 2 days depending on dose of Dox. Induced GFP mRNA expression returned to uninduced levels within 3 days upon Dox removal. Treatment of rtTA inducer and TRE responder double transgenic animals with Dox caused acceleration of metamorphic changes in thyroid hormone-response gene expression and morphology. These transgenic lines will be made available through the new Xenopus Stock Center and will serve as valuable tools for genetic analysis of development and metamorphosis.

Adipose tissue and skeletal and heart muscle, which exhibit insulin-stimulated glucose uptake, express a specific, insulin-responsive glucose transporter. Previously, a cDNA (GT2) encoding this protein was isolated from a mouse 3T3-L1 adipocyte library and was sequenced. Here we report the isolation and characterization of the corresponding mouse gene designated GLUT4. The GLUT4 gene spans 7 kilobases and consists of 11 exons and 10 introns. The start site of transcription was mapped 180 nucleotides upstream of the initial methionine codon. The GLUT4 promoter contains four potential binding sites for the nuclear transcription factor Sp1 as well as a CCAAT box. DNase I footprinting of the GLUT4 promoter with nuclear extracts from undifferentiated and differentiated 3T3-L1 cells revealed that a differentiation-specific nuclear factor binds in the region at position -258 relative to the start site of transcription. Purified CCAAT/enhancer binding protein (C/EBP) was found to bind at the same position. Transient cotransfection into 3T3-L1 preadipocytes of a GLUT4 promoter-

chloramphenicol acetyltransferase gene construct that contains the C/EBP binding site, together with a C/EBP expression vector, revealed that C/EBP trans-activates the GLUT4 promoter. We suggest that C/EBP plays an important role in tissue-specific, as well as metabolic, regulation of the insulin-responsive glucose transporter gene.

We studied the repair of UV- and aflatoxin B1 (AFB1)-induced damage in the human metallothionein (hMT) gene family. After exposure to either UV or AFB1, DNA damage was initially repaired faster in the DNA fragments containing the transcribed hMT-IA, hMT-IE, and hMT-IIA genes than in the genome overall. By 6 h posttreatment, there was at least twice as much repair in these genes as in the rest of the genome. Repair of UV damage in the hMT-IB gene, which shows cell-type specific expression, and in the hMT-IIB gene, which is a nontranscribed processed pseudogene, was about the same as in the rest of the genome, whereas repair of AFB1-induced damage was deficient in these two genes. Inducing transcription of the three expressed hMT genes with CdCl2 or of only the hMT-IIA gene with dexamethasone increased the initial rate of repair in the induced genes another twofold over the rate observed when they were transcribed at a basal level. The rates of repair in the hMT-IB and hMT-IIB genes were not altered by these inducing treatments. Transcription of the hMT genes was transiently inhibited after UV irradiation. Inducing transcription of the genes did not shorten this UV-induced delay. Thus, the efficiency of repair of damage in a DNA sequence is dependent on the level of transcriptional activity associated with that sequence. However, an increased efficiency in repair of a gene itself is not necessarily coupled to recovery of its transcription after DNA damage.

Stable reference genes are important for gene expression analyses such as quantitative PCR. The stability of 15 candidate reference genes that can be used to developing mouse gonads was thoroughly verified using combinations of multiple algorithms. The expression of these genes fluctuated greatly depending on the analysis period and/or gender. Peptidylprolyl isomerase A (Ppia) and polymerase (RNA) II (DNA directed) polypeptide A (Polr2a) were the reference genes that were used stably for a wide analysis period in developing mouse gonads. Furthermore, the stable reference genes corresponding to the analysis period and/or gender were ranked. These results are useful for the selection of the optimal reference gene required for high-precision measurements

Knowledge of an organism's fitness for survival is important for a complete understanding of microbial genetics and effective drug design. Current essential gene databases provide only binary essentiality data from genome-wide experiments. We therefore developed a new database that Integrates quantitative Fitness Information for Microbial genes (IFIM). The IFIM database currently contains data from 16 experiments and 2186 theoretical predictions. The highly significant correlation between the experiment-derived fitness data and our computational simulations demonstrated that the computer-generated predictions were often as reliable as the experimental data. The data in IFIM can be accessed easily, and the interface allows users to browse through the gene fitness information that it contains. IFIM is the first resource that allows easy access to fitness data of microbial genes. We believe this database will contribute to a better understanding of microbial genetics and will be useful in designing drugs to resist microbial pathogens, especially when experimental data are unavailable.

Mutations in oncogenes and tumor suppressor genes are responsible for tumorigenesis and represent favored therapeutic targets in oncology. We exploited homologous recombination to knock-in individual cancer mutations in the genome of nontransformed human cells. Sequential introduction of multiple mutations was also achieved, demonstrating the potential of this strategy to construct tumor progression models. Knock-in cells displayed allele-specific activation of signaling pathways and mutation-specific phenotypes different from those obtainable by ectopic oncogene expression. Profiling of a library of pharmacological agents on the mutated cells showed striking sensitivity or resistance phenotypes to pathway-targeted drugs, often matching those of tumor cells carrying equivalent cancer mutations. Thus, knock-in of single or multiple cancer alleles provides a pharmacogenomic platform for the rational design of targeted therapies.

In this paper, we extend our previous analyses of a set of genes in Caenorhabditis elegans that are involved in muscle structure and function: unc-93 III, sup-9 II, sup-10 X and sup-11 I. We describe an unusual, visible allele of sup-10, examine how this allele interacts genetically with mutations in other genes of this set and propose that the wild-type products of the unc-93 and sup-10 loci may be components of a protein complex. We also describe a new gene of this set, sup-18 III, and the interaction of sup-18 alleles with mutations in the other genes.

We report the complete molecular organization of the Dopa decarboxylase gene cluster. Mutagenesis screens recovered 77 new Df(2L)TW130 recessive lethal mutations. These new alleles combined with 263 previously isolated mutations in the cluster to define 18 essential genes. In addition, seven new deficiencies were isolated and characterized. Deficiency mapping, restriction fragment length polymorphism (RFLP) analysis and P-element-mediated germline transformation experiments determined the gene order for all 18 loci. Genomic and cDNA restriction endonuclease mapping, Northern blot analysis and DNA sequencing provided information on exact gene location, mRNA size and transcriptional direction for most of these loci. In addition, this analysis identified two transcription units that had not previously been identified by extensive mutagenesis screening. Most of the loci are contained within two dense subclusters. We discuss the effectiveness of mutagens and strategies used in our screens, the variable mutability of loci within the genome of Drosophila melanogaster, the cytological and molecular organization of the Ddc gene cluster, the validity of the one band-one gene hypothesis and a possible purpose for the clustering of genes in the Ddc region.

Induction of the adaptive response to alkylation damage results in the expression of four genes arranged in three transcriptional units: the ada-alkB operon and the alkA and aidB genes. Adaptive-response induction requires the ada gene product and occurs when cells are treated with methylating agents. In previous studies we noted that aidB, but not alkA or ada-alkB, was induced in the absence of alkylation damage as cells were grown to stationary phase. In this note we present evidence that aidB is induced by anaerobiosis. Thus, aidB is subject to dual regulation by ada-dependent alkylation induction and ada-independent anaerobic induction.

Mouse cells of the Mx+ genotype accumulate Mx mRNA in response to type I interferon (IFN). Nuclear runoff experiments show that IFN stringently regulates Mx gene expression at the level of transcription. Mx mRNA synthesis peaks about 3 h after IFN treatment, and within 5 h, Mx mRNA concentration rises from undetectable levels to about 0.1% of polyadenylated RNA.

The spread of a "cream" mutant in a wild population of house mice is reported. The hypothesis that the gene responsible for the colour, extreme chinchilla, ce, has spread because of linkage with a major gene for warfarin-resistance, is tested by a linkage backcross. The results prove that a major gene does exist, that it is very closely linked with frizzy, fr, in chromosome 7, which in turn is linked with ce, that it is fully dominant in females at 4 months of age, and that its partial dominance in males is under the control of modifiers. The symbol War is proposed for the gene. Its position in chromosome 7 is analagous with the position of the resistant gene, Rw2, in the rat in the analagous chromosome. The adaptive significance of the finding is discussed, as also are reports of certain other mutants in wild populations of mice.

Exogenous thiamine regulates Aspergillus oryzae thiA, which is involved in thiamine synthesis. One of the two introns in its 5'-untranslated region (5'-UTR) contains motifs (regions A and B) highly conserved among fungal thiamine biosynthesis genes. Deletion of either region relieved the repression by thiamine and thiamine inhibited intron splicing, suggesting that regions A and B are required for efficient splicing. Furthermore, transcript splicing was essential for thiA gene expression. These observations suggest a novel gene expression regulatory mechanism in filamentous fungi, in which exogenous thiamine controls intron splicing to regulate gene expression. Interestingly, regions A and B constitute a part of a thiamine pyrophosphate-binding riboswitch-like domain that has been quite recently found in the 5'-UTR of thiA.

Regulated antisense RNA (asRNA) expression has been employed successfully in Gram-positive bacteria for genome-wide essential gene identification and drug target determination. However, there have been no published reports describing the application of asRNA gene silencing for comprehensive analyses of essential genes in Gram-negative bacteria. In this study, we report the first genome-wide identification of asRNA constructs for essential genes in Escherichia coli. We screened 250 000 library transformants for conditional growth inhibitory recombinant clones from two shotgun genomic libraries of E. coli using a paired-termini expression vector (pHN678). After sequencing plasmid inserts of 675 confirmed inducer sensitive cell clones, we identified 152 separate asRNA constructs of which 134 inserts came from essential genes, while 18 originated from nonessential genes (but share operons with essential genes). Among the 79 individual essential genes silenced by these asRNA constructs, 61 genes (77%) engage in processes related to protein synthesis. The cell-based assays of an asRNA clone targeting fusA (encoding elongation factor G) showed that the induced cells were sensitized 12-fold to fusidic acid, a known specific inhibitor. Our results demonstrate the utility of the paired-termini expression vector and feasibility of large-scale gene silencing in E. coli using regulated asRNA expression.

Interferon-gamma (IFN-gamma) regulates a variety of immunoregulatory functions through the induction of a specific set of IFN-gamma response genes. This includes the invariant chain associated with the major histocompatibility complex class II molecules. To investigate the mechanism involved in the invariant chain (In) response to IFN-gamma we constructed chloramphenicol acetyltransferase (CAT) hybrid genes in which the CAT gene is under the control of the In promoter. The glioblastoma cell line, U-373 MG, transfected with a CAT construct having the In promoter sequence -790 to +1 bp showed over 3-fold increased CAT activity when treated with IFN-gamma indicating that this region confers IFN-gamma responsiveness to the CAT gene. The IFN-gamma response element in the promoter was further sublocalized to the region -120 to -61 base pairs (bp). This region contains homology to the interferon-stimulated response elements identified in other IFN responsive genes. By gel shift analyses, an IFN-gamma-induced sequence-specific DNA-binding factor was identified. This induced complex binds to an oligonucleotide corresponding to -107 to -79 bp of the In promoter. Mutations of this binding site at -94 and -92 bp drastically decreased binding of the constitutive and IFN-gamma-induced complexes. This IFN-gamma induced factor also binds to an oligonucleotide corresponding to -91 to -62 bp of the interferon-beta (IFN-beta) gene promoter, a region necessary for the induction of the IFN-beta gene by virus and double-stranded RNA. This binding specificity is characteristic of a family of DNA binding factors that bind both the interferon-stimulated response elements and the IFN-beta gene promoter.

We developed a high-copy-number plasmid system containing the entire structural and regulatory sequences of the phosphate-repressible acid phosphatase (PHO5) gene and the TRP1/ARS1 replicator sequences of the yeast Saccharomyces cerevisiae to investigate the mechanism of repression-derepression of transcription. The resulting plasmid was used to transform either wild-type cells or a number of strains which contain mutations in various trans-acting regulatory loci for the production of acid phosphatase. Results of analysis of mRNA levels isolated from the transformed strains grown under repressed or derepressed conditions suggested that normal transcriptional regulation of the gene persisted, although gene copy number was significantly increased. Analysis of changes in linking number (i.e., the number of negative supercoils) of the plasmid isolated under repressed and derepressed growth conditions revealed that the transcriptionally inactive plasmid contained approximately three more negative supercoils than the transcriptionally active plasmid. This difference in topological state was similarly seen in a plasmid containing a sequence-related acid phosphatase gene (PHO11) under the same regulatory control system, but it was not seen in plasmids isolated from some strains containing mutations which caused either fully constitutive or nonderepressible production of acid phosphatase. Finally, analysis of the nucleosome positioning along the inactive gene sequence revealed that an abnormally broad internucleosomal spacer is present in a region presumed to function in the regulation of transcription by the level of Pi in the growth media.

Transcription of the chicken ovalbumin gene is induced both in vivo and in vitro by four classes of steroid hormones. Recent experiments identified a steroid-dependent regulatory element (SDRE) in the 5'-flanking region of the ovalbumin gene between -900 and -521. To characterize the regulatory properties of the SDRE more precisely, additional mutations were created in this region, and fusion genes prepared by linking the ovalbumin 5'-flanking region and promoter to the chloramphenicol acetyltransferase structural gene. When the ovalbumin-chloramphenicol acetyltransferase fusion genes were transfected into steroid-responsive primary oviduct cells, mutants lacking sequences between -900 and -732 were no longer responsive to estrogen, corticosterone, progesterone, or dihydrotestosterone. The SDRE did not confer steroid-dependent expression on the heterologous thymidine kinase promoter by itself but did in conjunction with the negative regulatory element identified between -350 and -100. This suggests that the two elements act as a single functional entity and that the SDRE is not behaving as a typical steroid response element. Gel shift analyses revealed that two SDRE.protein complexes were formed when nuclear protein extracts were derived from estrogen-treated chicken oviduct but that only one complex was formed with extracts from estrogen-withdrawn oviduct or from other tissues. Neither an estrogen response element oligomer nor a glucocorticoid/progesterone response element oligomer competed for either of the DNA.protein complexes. Partially purified progesterone receptor also did not bind to the SDRE. These data indicate that induction of the ovalbumin gene by steroid hormones requires complex interactions involving both the SDRE and the negative regulatory element.

Basidiomycete fungi are an attractive resource for biologically active natural products for use in pharmaceutically relevant compounds. Recently, genome projects on mushroom fungi have provided a great deal of biosynthetic gene cluster information. However, functional analyses of the gene clusters for natural products were largely unexplored because of the difficulty of cDNA preparation and lack of gene manipulation tools for basidiomycete fungi. To develop a versatile host for basidiomycete genes, we examined gene expression using genomic DNA sequences in the robust ascomycete host *Aspergillus oryzae*, which is frequently used for the production of metabolites from filamentous fungi. Exhaustive expression of 30 terpene synthase genes from the basidiomycetes *Clitopilus pseudo-pinsitus* and *Stereum hirsutum* showed two splicing patterns, i.e., completely spliced cDNAs giving terpenes (15 cases) and mostly spliced cDNAs, indicating that *A. oryzae* correctly spliced most introns at the predicted positions and lengths. The mostly spliced cDNAs were expressed after PCR-based removal of introns, resulting in the successful production of terpenes (14 cases). During this study, we observed relatively frequent mispredictions in the automated program. Hence, the complementary use of *A. oryzae* expression and automated prediction will be a powerful tool for genome mining.

The human Fc gamma RI (CD64) is a high affinity receptor for the Fc portion of immunoglobulin (Ig), and its constitutively low expression on the cell surface of monocyte/macrophage and neutrophils is selectively upregulated by interferon gamma (IFN-gamma) treatment (Perussia, B., E. T. Dayton, R. Lazarus, V. Fanning, and G. Trinchieri. 1983. J. Exp. Med. 158:1092). Three distinct cDNAs have been cloned and code for proteins that predict three extracellular Ig-like domains (Allen, J.M., and B. Seed. 1989. Science [Wash. DC]. 243:378). Several differences in the coding region of these cDNAs suggest that in addition to polymorphic differences a second Fc gamma RI gene could possibly exist. This alternative Fc gamma RI gene (Fc gamma RIb) was defined by the lack of a genomic HindIII restriction site (van der Winkel, J. G. J., L. U. Ernst, C. L. Anderson, and I. M. Chiu. 1991. J. Biol. Chem. 266:13449). We describe the characterization a second gene (Fc gamma RIb) that has a termination codon in the third extracellular domain and therefore predicts a soluble form of a termination codon in the third extracellular domain and therefore predicts a soluble form of the receptor. We also define two distinct IFN-gamma-responsive regions in the 5' flanking sequence of Fc gamma RIb that resemble motifs that have been defined in the class II major histocompatibility complex promoter. The Fc gamma RIb promoter does not possess canonical TATA or CCAAT boxes, but does possess a palindromic motif that closely resembles the initiator sequence identified in the terminal deoxynucleotidyl transferase/human leukocyte IFN/adeno-associated virus type II P5 gene promoters (Smale, S. T., and

D. Baltimore. 1989. Cell. 57:103; Seto, E., Y. Shi, and T. Shenk. 1991. Nature [Lond.]. 354:241; Roy, A. L., M. Meisterernst, P. Pognonec, and R. C. Roeder. 1991. Nature [Lond.]. 354:245) virus type II P5 gene promoters raising interesting questions as to its role in the basal and myeloid-specific transcription of this gene.

Myxococcus xanthus is a soil-dwelling member of the δ-Proteobacteria that exhibits a complex developmental cycle upon starvation. Development comprises aggregation and differentiation into environmentally resistant myxospores in an environment that includes fluctuations in metal ion concentrations. While copper is essential for M. xanthus cells because several housekeeping enzymes use it as a cofactor, high copper concentrations are toxic. These opposing effects force cells to maintain a tight copper homeostasis. A plethora of paralogous genes involved in copper detoxification, all of which are differentially regulated, have been reported in M. xanthus. The use of in-frame deletion mutants and fusions with the reporter gene lacZ has allowed the identification of a two-component system, CorSR, that modulates the expression of an operon termed curA consisting of nine genes whose expression slowly increases after metal addition, reaching a plateau. Transcriptional regulation of this operon is complex because transcription can be initiated at different promoters and by different types of regulators. These genes confer copper tolerance during growth and development. Copper induces carotenoid production in a ΔcorSR mutant at lower concentrations than with the wild-type strain due to lack of expression of a gene product resembling subunit III of cbb3-type cytochrome c oxidase. This data may explain why copper induces carotenoid biosynthesis at suboptimal rather than optimal growth conditions in wild-type strains.

Positive control in vitro by gene N protein of bacteriophage lambda was demonstrated. lambda DNA was used to direct in vitro synthesis of lambda endolysin in a cell-free protein-synthesizing preparation derived from Escherichia coli. The endolysin synthesis depends on the concomitant in vitro synthesis of lambda gene N protein. When lambda N(-) DNA was used to direct the cell-free preparation, endolysin was made only if extract was added from cells in which a lambda prophage had been induced. The use of various prophage deletion strains proved that if this stimulating activity made in vivo is coded by a known lambda gene, it must be coded by gene N. The ability to stimulate endolysin synthesis in vitro on a lambda N(-) DNA template, therefore, constitutes an assay for N protein.

Enterococcus gallinarum N04-0414 (MIC for vancomycin, 256 microg/ml) harbored a vanD-type vancomycin resistance operon as well as the intrinsic vanC1 operon. The D-Ala:D-Ala ligase 2 gene (ddl2) was not present in the strain, though it is found downstream of the vanS gene from the vanC operon in E. gallinarum ATCC 49573 and 19 other E. gallinarum strains tested.

The recently discovered p53-related genes, p73 and p63, express multiple splice variants and N-terminally truncated forms initiated from an alternative promoter in intron 3. To date, no alternative promoter and multiple splice variants have been described for the p53 gene. In this study, we show that p53 has a gene structure similar to the p73 and p63 genes. The human p53 gene contains an alternative promoter and transcribes multiple splice variants. We show that p53 variants are expressed in normal human tissue in a tissue-dependent manner. We determine that the alternative promoter is conserved through evolution from Drosophila to man, suggesting that the p53 family gene structure plays an essential role in the multiple activities of the p53 family members. Consistent with this hypothesis, p53 variants are differentially expressed in human breast tumors compared with normal breast tissue. We establish that p53beta can bind differentially to promoters and can enhance p53 target gene expression in a promoter-dependent manner, while Delta133p53 is dominant-negative toward full-length p53, inhibiting p53-mediated apoptosis. The differential expression of the p53 isoforms in human tumors may explain the difficulties in linking p53 status to the biological properties and drug sensitivity of human cancer.

The calcium ionophore A23187 can reversibly induce the expression of two glucose-regulated genes, p3C5 and p4A3. This induction requires a continuous presence of the ionophore for over 2 h. Although

extracellular Ca2+ is important for the optimal effect of A23187, it is not necessary for the induction, since a similar response with a lower magnitude can be triggered in cells cultured in low Ca2+ medium buffered with EGTA. Both the basal and induced levels of p3C5 and p4A3 transcripts can be modulated by the calmodulin antagonist W-7, indicating the involvement of Ca2+/calmodulin-associated pathways. In addition, the sensitivity of the A23187 induction to cycloheximide suggests that the induction process is dependent on de novo protein synthesis.

At least two genes encode isoenzymes of rat 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Alternative splicing of one of these genes generates a skeletal muscle-specific transcript from an upstream promoter and a liver-specific transcript from a downstream promoter. A potent glucocorticoid response element was identified in the first intron of the gene, i.e. between liver exon I and exon II. The element is approximately 3.5 kilobase pairs (kb) downstream of the liver isoenzyme transcription start site and 13 kb upstream of exon II of the gene and confers dexamethasone-sensitive expression of chloramphenicol acetyltransferase (CAT) activity from a heterologous thymidine kinase promoter and from both homologous 5'-flanking regions of the gene. This glucocorticoid response element also exhibits androgen- but not estrogen-sensitive expression of CAT activity in HeLa cells cotransfected with the appropriate receptor expression vector. DNase footprint and sequence analysis revealed that the element is comprised minimally of two adjacent 15-mer glucocorticoid receptor dimer binding sites situated in opposite orientations. Glucocortcoid regulation of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase gene expression in liver and skeletal muscle is mediated by a single complex glucocorticoid response element located in the first intron of the skeletal muscle/liver gene.

CXCR4 is the receptor for the CXC chemokine SDF1 that has essential functions on embryo organogenesis, immunological functions and T lymphocyte trafficking. Recently, CXCR4 has drawn unexpected attention as it was recently identified as a co-factor required for entry of lymphotropic HIV isolates in CD4+ T lymphocytes. CXCR4 is the only SDF1 receptor identified so far. This suggests that CXCR4 expression is critical for the biological effects of SDF1. To investigate the mechanisms controlling both the constitutive and induced expression of CXCR4 receptors we have isolated and characterized the promoter region and determined the genomic structure of the human gene. The CXCR4 gene contains two exons separated by an intronic sequence. A 2.6 kb 5'-flanking region located upstream the CXCR4 open reading frame contains a TATA box and the transcription start site characteristic of a functional promoter. This region also contains putative consensus binding sequences for different transcription factors, some of them associated with the hemopoiesis and lymphocyte development.

The immunosuppressive drug tacrolimus, whose pharmacokinetic characteristics display large interindividual variations, is a substrate for P-glycoprotein (P-gp), the product of the multidrug resistance-1 (MDR1) gene. Some of the single nucleotide polymorphisms (SNP) of MDR1 reported correlated with the in vivo activity of P-gp. Because P-gp is known to control tacrolimus intestinal absorption, it was postulated that these polymorphisms are associated with tacrolimus pharmacokinetic variations in renal transplant recipients. The objective of this study was to evaluate in a retrospective study of 81 renal transplant recipients the effect on tacrolimus dosages and concentration/dose ratio of four frequent MDR1 SNP possibly associated with P-gp function (T-129C in exon 1b, 1236C>T in exon 12, 2677G>T,A in exon 21, and 3435C>T in exon 26). As in the general population, the SNP in exons 12, 21, and 26 were frequent (16, 17.3, and 22.2% for the variant homozygous genotype, respectively) and exhibited incomplete linkage disequilibrium. One month after tacrolimus introduction, exon 21 SNP correlated significantly with the daily tacrolimus dose ($P < or = 0.05$) and the concentration/dose ratio ($P < or = 0.02$). Tacrolimus dose requirements were 40% higher in homozygous than wild-type patients for this SNP. The concentration/dose ratio was 36% lower in the wild-type patients, suggesting that, for a given dose, their tacrolimus blood concentration is lower. Haplotype analysis substantiated these results and suggested that exons 26 and 21 SNP may be associated with tacrolimus dose requirements. Genotype monitoring of the MDR1 gene reliably predicts the optimal dose of tacrolimus

in renal transplant recipients and may predict the initial daily dose needed by individual patients to obtain adequate immunosuppression.

Group I catalytic introns are widespread in bacterial, archaeal, viral, organellar, and some eukaryotic genomes, where they are reported to provide regulatory functions. The group I introns are currently divided into five types (A-E), which are themselves distributed into several subtypes, with the exception of group I type D intron (GI-D). GI-D introns belong to the rarest group with only 17 described to date, including only one with a putative role reported in fungi, where it would interfere with an adaptive response in the cytochrome b (COB) gene to quinone outside inhibitor (QoI) fungicide resistance. Using homology search methods taking into account both conserved sequences and RNA secondary structures, we analysed the mitochondrial genomes or COB genes of 169 fungal species, including some frequently under QoI selection pressure. These analyses have led to the identification of 216 novel GI-D introns, and the definition of three distinct subtypes, one of which being linked with a functional activity. We have further uncovered a homing site for this GI-D intron type, which helps refine the accepted model of quinone outside inhibitor resistance, whereby mobility of the intron across fungal mitochondrial genomes, would influence a fungus ability to develop resistance to QoIs.

Afobazole, a new 2-mercapto-benzimidazole derivative, exhibited antimutagenic activity in chromosome aberration tests and antioxidant properties. The aim of this study was to demonstrate the potential chemopreventive effect of afobazole on the level of early biological effects by analysing changes in oncogene and tumor suppressor gene expression.

Amplification of the M2 gene encoding the small subunit of ribonucleotide reductase (EC 1.17.4.1) was analyzed in a collection of vaccinia virus (VV) isolates selected for resistance to 5 mM hydroxyurea (HU). Most of the mutants harbored tandem direct repeat arrays of the M2 gene, but several had duplicated M2 as an inverted repeat by genomic rearrangements involving the chromosomal termini. Novel joints formed by direct repeats were mapped, amplified in vitro, and sequenced. The junctions were simple fusions between DNA downstream and upstream of the M2 gene. Lack of sequence homology at the breakpoints indicated that the initial genomic rearrangements leading to gene amplification were due to nonhomologous recombination events.

In Saccharomyces cerevisiae the anaerobic (oxygen-repressed) ANB1 gene and a group of aerobic (oxygen-induced) genes are coordinately regulated by the ROX1 gene. We report here that heme, known as an inducer of aerobic genes, also causes inhibition of ANB1 expression. Thus, in combination with the ROX1 gene product heme has an opposite effect on the expression of anaerobic and aerobic genes. Accumulation of ANB1 mRNA was sharply decreased in anaerobic cells grown in the presence of heme. This effect must operate at the level of transcription since heme also inhibited accumulation of CYC1 mRNA from an ANB1-CYC1 fusion. Heme precursors did not appear to function either as inhibitors or as activators. Oxygen itself also had no effect on transcription of ANB1. Repression by heme cannot be attributed to the respiratory competence conferred by heme since both ANB1 and the aerobic genes tr-1 and CYC1 were regulated normally in [rho 0] mutants. The results are consistent with a classical allosteric coeffector function for heme, although more indirect explanations are tenable. A role for the ROX1 gene product in transcriptional regulation can be inferred from the observation that there was no inhibition of ANB1 expression by heme in rox1 mutants. Judging from this epistasis the rox1 phenotype is not due to a defect in heme production; this would indicate that the ROX1 factor functions by mediating the effect of heme on transcription.

We describe pOp/LhGR, a dexamethasone-inducible derivative of the pOp/LhG4 transcription activation system, and its use in tobacco to regulate expression of uidA (encoding beta-glucuronidase; GUS) and the cytokinin-biosnythetic gene ipt. The pOp/LhGR system exhibited stringent regulation and strong induced phenotypes in soil and tissue culture. In conjunction with an improved target promoter, pOp6, that carries six copies of an optimized lac operator sequence the pOp6/LhGR system directed induced GUS activities that exceeded those obtained with pOp/LhG4 or the CaMV 35S

promoter but without increased uninduced activity. A single dose of dexamethasone was sufficient to direct cytotoxic levels of ipt expression in soil-grown plants although uninduced plants grew normally throughout a complete life cycle. In vitro, induced transcripts were detectable within an hour of dexamethasone application and 1 nM dexamethasone was sufficient for half maximal induction of GUS activity. Various methods of dexamethasone application were successfully applied under tissue culture and greenhouse conditions. We observed no inhibitory effects of dexamethasone or LhGR on plant development even with the highest concentrations of inducer, although tobacco seedlings were adversely affected by ethanol used as a solvent for dexamethasone stock solutions. The pOp/LhGR system provides a highly sensitive, efficient, and tightly regulated chemically inducible transgene expression system for tobacco plants.

Defective tumor suppressor genes (TSGs) and hyperactive oncogenes (OCGs) heavily contribute to cell proliferation and apoptosis during cancer development through genetic variations such as somatic mutations and deletions. Moreover, they usually do not perform their cellular functions individually but rather execute jointly. Therefore, a comprehensive comparison of their mutation patterns and network properties may provide a deeper understanding of their roles in the cancer development and provide some clues for identification of novel targets.

Ustiloxins were found recently to be the first example of cyclic peptidyl secondary metabolites that are ribosomally synthesized in filamentous fungi. In this work, two function-unknown genes (ustYa/ustYb) in the gene cluster for ustiloxins from Aspergillus flavus were found experimentally to be involved in cyclization of the peptide. Their homologous genes are observed mainly in filamentous fungi and mushrooms. They have two "HXXHC" motifs that might form active sites. Computational genome analyses showed that these genes are frequently located near candidate genes for ribosomal peptide precursors, which have signal peptides at the N-termini and repeated sequences with core peptides for the cyclic portions, in the genomes of filamentous fungi, particularly Aspergilli, as observed in the ustiloxin gene cluster. Based on the combination of the ustYa/ustYb homologous genes and the nearby ribosomal peptide precursor candidate genes, 94 ribosomal peptide precursor candidates that were identified computationally from Aspergilli genome sequences were classified into more than 40 types including a wide variety of core peptide sequences. A set of the predicted ribosomal peptide biosynthetic genes was experimentally verified to synthesize a new cyclic peptide compound, designated as asperipin-2a, which comprises the amino acid sequence in the corresponding precursor gene, distinct from the ustiloxin precursors.

A method for transforming Tetrahymena has been established earlier, but its application has been limited because of the lack of selectable markers other than the rRNA-encoding DNA (rDNA). Mutations in the yeast ribosomal protein L29 gene (CYH2) are known that confer cycloheximide resistance. We have cloned and sequenced the homologue of this gene from both a wild-type and a cycloheximide-resistant (ChxA) strain of Tetrahymena. Surprisingly, a comparison shows that the ChxA mutation is not present in the CYH2 homologue. We therefore created the yeast mutations in the Tetrahymena gene by site-directed mutagenesis and used them to transform Tetrahymena either with or without linking to an rDNA vector. All clones transformed by the rDNA vector also became resistant to cycloheximide when the rDNA contained the engineered mutant genes. Without the rDNA vector, the mutant genes transform approximately 1% of injected cells to become resistant to cycloheximide. DNA analysis indicates that transformation occurs by replacement of the host sequence and not by random integration of the injected sequence. The replacement occurs to some but not all copies of this gene in the polyploid macronuclear genome. Thus, transformation in Tetrahymena occurs by specific sequence replacement, and the injected mutant genes can serve as dominant selectable transformation markers in this organism.

Transcription of the vasoactive intestinal polypeptide (VIP) gene is regulated by cAMP. To identify the nucleotide sequences in the human VIP gene responsible for this regulation, we constructed chimeric genes containing different portions of the 5'-flanking region of the human VIP gene fused to the

structural sequence encoding the bacterial reporter enzyme chloramphenicol acetyltransferase (CAT). The transcriptional activities of the fusion genes introduced into the rat pheochromocytoma cell line PC12 were assayed by measuring CAT activity in the cell lysates. Forskolin, an adenylate cyclase-activating agent, stimulated the expression of VIP-CAT fusion genes. Deletional analysis demonstrated that a region between -86 and -70 nucleotides upstream from the transcriptional origin of the human VIP gene was responsible for stimulation by forskolin. This region was able to confer cAMP-responsiveness to a gene that is not normally regulated by cAMP. Two copies of a 5 base pair motif, 5'-CGTCA-3', are required for activity of the VIP cAMP regulatory region. This motif is also present in the cAMP regulatory region of several other eukaryotic genes.

The expression of cellobiohydrolase I mRNA from Trichoderma reesei, measured by Northern blot hybridization, is controlled by the nature of carbon sources used in the culture medium. Cellulose and the soluble disaccharide sophorose, but not glycerol or glucose, act as inducers. Cellobiohydrolase I mRNA was undetectable when antibodies to the major members of the cellulolytic system were present in the culture medium prior to the addition of cellulose. These antibodies had no repressive effect if sophorose was used as an inducer. The results strongly suggest that a low constitutive cellulolytic system catalyzes the formation of a soluble inducer from cellulose and that this inducer triggers the expression of the cellobiohydrolase I gene transcript, most probably at the transcription level.

The genetic determinants of enterobacterial common antigen (ECA) include the rfe and rff genes located between ilv and cya near min 85 on the Escherichia coli chromosome. The rfe-rff gene cluster of E. coli K-12 was cloned in the cosmid pHC79. The cosmid clone complemented mutants defective in the synthesis of ECA due to lesions in the rfe, rffE, rffD, rffA, rffC, rffT, and rffM genes. Restriction endonuclease mapping combined with complementation studies of the original cosmid clone and six subclones revealed the order of genes in this region to be rfe-rffD/rffE-rffA/rffC-rffT-rffM . The rfe gene was localized to a 2.54-kilobase ClaI fragment of DNA, and the complete nucleotide sequence of this fragment was determined. The nucleotide sequencing data revealed two open reading frames, ORF-1 and ORF-2, located on the same strand of DNA. The putative initiation codon of ORF-1 was found to be 570 nucleotides downstream from the termination codon of rho. ORF-1 and ORF-2 specify putative proteins of 257 and 348 amino acids with calculated Mr values of 29,010 and 39,771, respectively. ORF-1 was identified as the rfe gene since ORF-1 alone was able to complement defects in the synthesis of ECA and 08-side chain synthesis in rfe mutants of E. coli. Data are also presented which suggest the possibility that the rfe gene is the structural gene for the tunicamycin sensitive UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase that catalyzes the synthesis of GlcNAc-pyrophosphorylundecaprenol (lipid I), the first lipid-linked intermediate involved in ECA synthesis.

A group of five cDNA clones, representing the gadd genes, were recently isolated from Chinese hamster ovary (CHO) cells as genes induced upon growth arrest and after DNA damage (Fornace, A. J., Jr., Nebert, D. W., Hollander, M. C., Luethy, J. D., Papathanasiou, M., Fargnoli, J., and Holbrook, N. J. (1989) Mol. Cell. Biol. 9, 4196-4203). We have isolated and characterized one of these genes, gadd153. The gene spans five kilobases and contains four exons. The 5'-flanking region of the gene, within 420 base pairs of the transcription initiation site, contains a number of cis elements associated with transcriptional regulation in other genes. These include a Hogness box, ATAAAA, an inverted GCCAAT box; seven SP1 transcription factor binding sites, and an AP-1 site. This region is rich in G + C content (greater than 70%) and contains an unusually long stretch of alternating CpG residues. The 800-base pair region immediately upstream of the transcription start site can drive expression of the bacterial chloramphenicol acetyltransferase (CAT) gene, but only in its endogenous orientation, in three different cell lines: HeLa, CHO, and Jurkat. The gadd153 promoter is strongly activated by methyl methanesulfonate, hydrogen peroxide, and UV irradiation, but not by growth arrest signals. This suggests that separate and very different regulatory pathways are involved in the induction of the gadd153 gene by growth cessation and DNA damage.

Suicide genes have been widely investigated for their utility as therapeutic agents and as tools for in vitro negative selection strategies. Several methods for delivery of suicide genes have been explored. Two important considerations for delivery are the quantity of delivered cargo and the ability to target the cargo to specific cells. Delivery using a lentiviral vector is particularly attractive due to the ability to encode the gene within the viral genome, as well as the ability to limit off-target effects by using cell type-specific glycoproteins. Here, we present the design and validation of a diphtheria toxin A (DTA)-encoding lentiviral vector expressing DTA under the control of a constituitive promoter to allow for expression of DTA in a variety of cell types, with specificity provided via selection of glycoproteins for pseudotyping of the lentiviral particles. DTA exerts its toxic activity through inhibition of eukaryotic translation elongation factor 2 (eEF2) via adenosine diphosphate (ADP)-ribosylation of a modified histidine residue, diphthamide, at His715, which blocks protein translation and leads to cell death. Thus, we also detail development of DTA-resistant cell lines, engineered through CRISPR/Cas9-mediated knockout of the diphthamide 1 (DPH1) gene, which enable both robust virus production by transfection and evaluation of DTA-expressing virus infectivity.

We have studied the transcriptional regulation of the beta 3 tubulin gene by the steroid hormone 20-hydroxyecdysone (20-OH-E) in Drosophila Kc cells. A series of hybrid genes with varying tubulin gene lengths driving the bacterial chloramphenicol acetyl transferase (CAT) gene were constructed. The promoter activity was assayed after transient expression in Kc cells, in the presence or absence of 20-OH-E. We find that 0.91Kb upstream from the transcription start site contain one or several hormone independent positive cis-acting elements, responsible for the constitutive expression of the beta 3 tubulin gene. In the large (4.5 Kb) first intron of this gene, we identified additional hormone dependent negative and positive regulatory elements, which can act in both directions and in a position-independence manner. Then, the negative intron element(s), which repress the transcription in the absence of 20-OH-E has characteristics of silencer.

To elucidate the mechanism by which abscisic acid (ABA) regulates gene expression, the promoter of the barley ABA-responsive HVA22 gene has been analyzed by both loss- and gain-of-function studies. Previous reports indicate that G-box sequences, which are present in genes responding to a variety of environmental and physiological cues, are involved in ABA response. However, our data suggest that G-box sequences are necessary but not sufficient for ABA response. Instead, an ABA response complex consisting of a G-box, namely, ABRE3 (GCCACGTACA), and a novel coupling element, CE1 (TGCCACCGG), is sufficient for high-level ABA induction, and replacement of either of these sequences abolishes ABA responsiveness. We suggest that the interaction between G-box sequences, such as ABRE3 in the HVA22 gene, and CE-type sequences determines the specificity in ABA-regulated gene expression. Our results also demonstrate that the ABA response complex is the minimal promoter unit governing high-level ABA induction; four copies of this 49-bp-long complex linked to a minimal promoter can confer more than 100-fold ABA-induced gene expression. In addition to ABA response complex 1, composed of ABRE3 and CE1, the HVA22 promoter contains another ABA response complex. The ABA responsiveness of this ABA response complex 2 relies on the interaction of G-box (ABRE2; CGCACGTGTC) with another yet unidentified coupling element. These two complexes contribute incrementally to the expression level of HVA22 in response to ABA.

We have isolated, mapped and sequenced adipsin, the adipocyte differentiation-dependent serine protease gene. This gene, which is present in a single form in the mouse, spans 1.7 kilobases and contains five exons. While the basic exon structure characteristic of serine protease genes is conserved in adipsin, there is also a fusion of two exons that are separate in other serine proteases. The sequence data also suggests a mechanism of alternative splicing which appears to account for the generation of two adipsin mRNA species differing by only three nucleotides and encoding two different signal peptides. To investigate the control of adipsin expression we have examined the effects of tumor necrosis factor (TNF) on adipocytes. The level of adipsin RNA is dramatically decreased by hormone treatment, but the change occurs more slowly than for other fat cell mRNAs, such as glycerophosphate dehydrogenase. These results show that adipsin is a novel serine protease gene whose expression is

regulated by a macrophage-derived factor which modulates expression of other adipocyte-specific RNAs.

The primary virulence determinant of Plasmodium falciparum malaria parasite-infected cells is a family of heterogeneous surface receptors collectively referred to as PfEMP1. These proteins are encoded by a large, polymorphic gene family called var. The family contains approximately 60 individual genes, which are subject to strict, mutually exclusive expression, with the single expressed var gene determining the antigenic, cytoadherent, and virulence phenotype of the infected cell. The mutually exclusive expression pattern of var genes is imperative for the parasite's ability to evade the host's immune response and is similar to the process of "allelic exclusion" described for mammalian Ig and odorant receptor genes. In mammalian systems, mutually exclusive expression is ensured by negative feedback inhibition mediated by production of a functional protein. To investigate how expression of the var gene family is regulated, we have created transgenic lines of parasites in which expression of individual var loci can be manipulated. Here we show that no such negative feedback system exists in P. falciparum and that this process is dependent solely on the transcriptional regulatory elements immediately adjacent to each gene. Transgenic parasites that are selected to express a var gene in which the PfEMP1 coding region has been replaced by a drug-selectable marker silence all other var genes in the genome, thus effectively knocking out all PfEMP1 expression and indicating that the modified gene is still recognized as a member of the var gene family. Mutually exclusive expression in P. falciparum is therefore regulated exclusively at the level of transcription, and a functional PfEMP1 protein is not necessary for viability or for proper gene regulation in cultured parasites.

All available amino acid-requiring mutants of Aspergillus nidulans were found to be hypersensitive to MMS (methyl methanesulfonate) to various degrees. On MMS media, secondary mutations could be selected which suppress this MMS sensitivity but do not affect the requirement. Many such mutations were analyzed and found to be alleles of one gene, smsA (= suppressor of MMS sensitivity), which mapped distal on the right arm of chromosome V. This gene is more likely to be involved in general regulation of amino acid biosynthesis than MMS uptake, since a variety of pathway interactions were clearly modified by smsA suppressors in the absence of MMS.

The regulatory properties of mouse pancreatic amylase genes include exclusive expression in the acinar cells of the pancreas and dependence on insulin and glucocorticoids for maximal expression. We have characterized a murine pancreatic amylase gene, Amy-2.2y, whose promoter sequence is 30% divergent from those of previously sequenced amylase genes. To localize sequences required for tissue-specific and hormone-dependent activation, we established two lines of transgenic mice. The first line contained a single copy of the complete Amy-2.2y gene as well as 9 kilobases of 5'-flanking sequence and 5 kilobases of 3'-flanking sequence. The second line carried a minigene which included 208 base pairs of 5'-flanking sequence and 300 base pairs of 3'-flanking sequence. In both lines the transgene was expressed at high levels exclusively in the pancreas. Both constructs were dependent on insulin and induced by dexamethasone. Thus, the transferred genes contained the sequences required for tissue-specific and hormonally regulated expression.