Functional genomics in virology and antiviral drug discovery

Victor DeFilippis, Camilo Raggo, Ashlee Moses and Klaus Früh

Virology research and antiviral drug discovery are poised to benefit from the post-genomic revolution for three main reasons. First, viruses need the host to replicate and are therefore vulnerable to inhibition of cellular pathways. Knowledge of complete genomic sequences of both virus and host now permits the study of this interplay on a global scale. Combining transcriptomics and proteomics with large-scale gene knockdown experiments will enable the identification of novel antiviral targets. Second, massive parallel assay systems, such as DNA microarrays, which define the post-genomic era, will facilitate viral diagnostics. Third, the combination of genetics with genomics will enable the analysis of viral mutants and strains on an unprecedented scale. The dramatic effects of viral infection on host cell transcriptional patterns have been well-documented and will be briefly highlighted. In addition, we discuss recent trends that apply functional genomics methods to the discovery of new targets and therapies for viral disease.

Virology has been a genomic science for longer than any other field because the first complete genomes sequenced were of viral origin [1]. As of this publication, 1535 full viral genome sequences are available in GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html). However, only modern sequencing endeavors made possible the complete sequencing of the human genome as well as that of model organisms. ‘Virogenomics’ research aims to describe the interaction between the products of two genomes using post-genomic methods such as DNA microarrays. Given that viruses require a host organism to propagate, this approach promises to deepen our understanding of virus–host interactions. Eventually, virogenomics research will also lead to new treatments for diseases caused by viruses, because whole genome sequences (virus and host) represent the complete collection of possible antiviral targets. Classically, antiviral drug discovery focused on viral gene products. However, the limited number of viral targets and the rapid emergence of drug-resistant mutations in many viral systems make the search for host cell targets an attractive alternative.

Modern virology and antiviral drug discovery are thus expected to be impacted dramatically by functional genomics methods. The widespread use of these methods is reflected in issues of major virology journals, which are rarely published without articles that include or rely on functional genomics experiments, in particular, DNA microarrays. However, most of the published studies are limited to showing that a given virus induces or represses a given set of genes. The future challenge, however, is to determine which of the myriad of host cell gene products are essential for virus survival or for virally induced disease progression. Such gene products represent targets for inhibiting virus infection, pathogenesis, or virus-induced tumor formation. Recent breakthroughs in gene knockdown technology are expected to accelerate the pace by which host cell pathways crucial to virus replication can be identified. Furthermore, diagnostic virology will be impacted by microarray technology because probes for thousands of viral species and strains can be matched against an unknown specimen. Here, we highlight some published and as-yet unpublished studies that use DNA microarrays and gene knockdown methods in virology.

Studying viral transcriptomes with DNA microarrays

A thorough discussion of the widespread use of microarray technology in biomedicine is beyond the scope of this review, but there are many excellent papers that cover these subjects in depth [2–4]. Briefly, microarrays are glass, silicon, or nylon platforms containing oligo- or polynucleotides (targets) identical or complementary to known genes. Interrogation of biological samples is performed by synthesis from sample mRNA of fluorescently labeled probes that are hybridized to target sequences. Gene expression is quantified by measuring fluorescence emission with a confocal laser scanner. The ability to immobilize up to 20 000 different genes on a single array enables examination of the transcriptional activity of entire eukaryotic genomes in parallel. However, microarrays are adaptable because the investigator can specify target sequences. Examination of viral genes, host genes or any such combination can thus be accomplished on a single array. This technology can also be used to immobilize specific probes for many different viral species, different viral strains, or complex DNA viruses.

Many common viruses, such as retroviruses, papillomaviruses and paroviruses, have very small genomes (often <10 genes). Therefore, using microarrays for examination of gene expression in such viruses is not practical. In these cases other techniques, such as northern blotting and quantitative RT-PCR (qPCR), are more appropriate. However, microarrays are useful for genotyping small viruses, characterization of strain variability
and identification of virus type(s) present in a clinical sample. High-density oligonucleotide arrays were used to characterize sequence variability of the HIV quasi-species in patient samples [5] and oligonucleotide arrays were applied to correlate progression of cervical cancer with the presence of the HPV genotype [6]. In a particularly clever study, Joe De Risi’s laboratory [7] constructed microarrays using polynucleotides of conserved regions from a broad range of common viruses. Using random PCR amplification from infected samples, they detected infecting viruses by examining specific patterns of hybridization. Recently, this technique confirmed that the agent of severe acute respiratory syndrome (SARS) is a member of the coronavirus family (CDC press release, 24 March 2003). Samples from SARS patients reacted with probes from avian and bovine coronaviruses. Microarrays allow not only simultaneous diagnosis of multi-strain viral infection from clinical specimens but also identification of new virus types (see [8,9]). Furthermore, diagnostic microarrays will have important applications in the fields of biological warfare and bioterrorism. DNA microarrays will be able to rapidly diagnose individuals for exposure to different biothreats such as poxviruses, plague, anthrax and tularemia. These methods can also be used to determine whether viral strains and types used in an attack have been genetically altered.

Viral genome microarrays are useful for determining the complex transcriptional program of larger viruses, especially members of the herpesvirus family. For instance, the transcriptional program of the 250 kilobase human cytomegalovirus (HCMV) genome [10] was mapped by Chambers and colleagues using oligonucleotide arrays representing all 226 HCMV open reading frames (ORFs) [11]. Because the function and expression pattern of many of the HCMV genes are still unknown, such classification might help in characterizing unknown ORFs (e.g. non-structural proteins are generally immediate early and early genes whereas structural proteins are often late genes).

Transcriptional analysis of the tumorigenic y-2 herpesvirus Kaposi’s sarcoma-associated virus (KSHV, human herpesvirus type 8) has also been examined with microarrays. The virus is mostly latent in infected and transformed cells but lytic replication can be induced either spontaneously in a small percentage of cells or artificially using phorbol esters in culture. Correlation of viral gene expression with latent and lytic phenotypes has been investigated using nylon membrane [12] and glass slide-based microarrays [13]. Jenner and colleagues [12] measured KSHV transcription in latently infected cell cultures that were both lysis-induced and uninduced over a time course (0–72 hr). Expression levels for each gene were grouped according to similarity in patterns of expression over all time points and both treatments by cluster analysis [14]. Interestingly, this approach showed that some genes clustering with typical lytic genes had been previously misclassified as latent using other methods. This was because their expression level was very high in a small percentage of cells undergoing spontaneous lytic replication thereby contaminating the ‘strictly latent’ sample. Other large DNA viruses, such as poxviruses, are also excellent candidates for such research. The emergence of smallpox as a potential agent of bioterrorism and recent monkeypox outbreaks have renewed interest in finding new therapeutic, diagnostic and immunological targets for improved diagnosis, treatment and vaccination strategies for poxviruses.

Viral genome arrays also allow phenotypic classification of viral mutants. For instance, gene expression was examined using microarrays for herpes simplex virus type 1 (HSV1) employing both wild type and a mutant strain lacking immediate-early gene α27 [15]. Although the mutant used in this study was well characterized, it is conceivable that unknown viral mutants can be classified according to their transcriptome fingerprint in a manner similar to the compendium approach described for yeast [16]. Particularly appealing is use of this approach to characterize and classify randomly generated mutants with targeted mutants. Much progress has been made in creating random transposon libraries of herpesvirus genomes cloned as bacterial artificial chromosomes (BACs) [17]. Transcriptional profiles observed in the HCMV and HSV1 experiments agreed strongly with previous studies of gene expression that used alternative techniques (e.g. northern blotting, RNAse protection assays, primer extension) thus validating the use of microarrays for examining gene transcription.

DNA microarray analysis of virally infected host cells

A complementary approach to viral mutant characterization is study of their role in manipulating the host cell transcriptional profile. The deletion of non-essential genes often results in a ‘normal’ phenotype in tissue culture – that is, the mutant virus behaves similar to the wild type in traditional assays such as single-step and multi-step growth curves. However, the microarray fingerprint of such a mutant might be very different from the wild type because the deleted gene might be involved in manipulating host cell pathways that are not essential in vitro but have an important role in vivo. An example for such a situation from our laboratory is shown in Fig. 1. A mutant CMV lacking a nonessential gene grew much like wild type in fibroblasts. However, microarray analysis revealed dramatic differences in the transcriptional profile of fibroblasts infected with the mutant versus wild-type virus. In addition to regulating a similar set of genes as the wild type, the mutant virus induced or repressed a unique set of genes not regulated in cells infected with the wild type. Such results will not only help to elucidate the function of nonessential viral genes but also could be used to classify viral mutants generated by random approaches.

The investigation of host cell transcriptional changes in response to virus infection and viral protein expression has been a vigorous area of microarray research over the past few years. At last count (July 2003) there are 201 references containing the words ‘microarray’ and ‘virus’ in the PubMed literature database, most of which describe the host cell response to viral infection. Given that virus replication involves use and manipulation of multiple host proteins and molecular processes, cellular pathways related to transcription and translation, signal transduction, metabolism, host defense and cell cycle control are
commonly altered in response to, or as a result of, infection with diverse virus types. Transcriptional changes in virally infected cells are either the result of anti-viral, 'pro'-viral or bystander host responses. As a common anti-viral response, expression of interferon-stimulated genes (ISGs) is often observed during infection by diverse, unrelated viruses. ISGs create a cellular state in which virus replication is blocked and thus there is natural selective pressure to circumvent or impair this host response (for a recent review see [18]). The ability to prevent, inhibit, or limit this induction might therefore contribute to virulence and represent a target pathway for antiviral treatments. For anti-viral therapy, it is conceivable to target viral proteins that interfere with the induction of ISGs thus releasing the innate immune response to combat viral infection.

An anti-host strategy used by viruses that complicates transcriptome analysis involves interference with cellular RNA metabolism. Global inhibition of transcription or RNA degradation can appear as gene 'repression' in a comparative analysis. A bias toward gene downregulation was indeed observed for influenza as well as HSV [15]. Interestingly, in each case several host transcripts were upregulated despite massive nonspecific downregulation of host cell mRNA. Picornaviruses induce degradation of the mRNA cap binding complex eIF4F thereby inhibiting cellular gene expression by preventing recruitment of 40S ribosomal subunits to 5' caps. Following this, translation of genes can only occur when they contain internal ribosome entry sites (IRES), as do many viral genes. By selectively isolating mRNA-polysome complexes in poliovirus infected (eIF4F depleted) and uninfected cells, Johannes et al. subsequently used microarrays to identify cellular genes that were translated during poliovirus infection [20]. Indeed, a small group of cellular genes were identified as being transcribed despite the lack of 5' caps in infected cells. Although these genes had diverse functions it seems possible that the internal ribosomal entry site evolved to overcome the block by picornaviruses. Thus, microarray analysis of polysome-fractions can be used to screen for host cell mRNAs that survive viral attack thus potentially representing novel anti-viral host factors.

In contrast to host pathways that represent an antiviral response, others might be beneficial or even essential for viral replication. Inhibiting such 'pro'-viral factors offers novel avenues for anti-viral drug discovery. There are several examples for such essential host factors that were identified by traditional, hypothesis-driven methods (for recent reviews see [21,22]). In addition, non-hypothesis-driven global gene expression profiling can reveal host cell genes that promote viral growth. However, the challenge is how to pick those genes that might be important for viral replication. Several investigators, including our laboratory, chose certain microarray-identified differentially regulated genes because they were associated with interesting biological characteristics or displayed unusual expression patterns. For instance, T. Shenk's group noticed in a DNA microarray experiment that several genes regulating prostaglandin metabolism were induced by human cytomegalovirus [23]. One of the induced genes was cyclooxygenase-2, which leads to the synthesis of prostaglandin E2 (PGE2) thus modulating cellular gene expression and immune function. Based on this initial observation, Zhu et al. recently demonstrated that HCMV replication requires the function of cyclooxygenase-2 [24]. The authors showed how specific inhibitors of COX-2 lead to a >100-fold decrease in yield of HCMV titers. Furthermore, the addition of PGE2 in the presence of such inhibitors restored normal HCMV replication. Intriguingly, a virus closely related to HCMV that infects rhesus macaques actually carries a COX-2 homologue in its genome [25], which further emphasizes the importance of this host cell gene for cytomegalovirus reproduction.

A recent example from Jay Nelson's laboratory further supports the notion that some transcriptionally upregulated host cell genes are important for viral replication (A. Hirsch and J. Nelson, pers. commun.). Examination of host gene expression in human cells infected with various flaviviruses revealed significant upregulation of a host cell kinase. Cells treated with corresponding kinase inhibitors or with small interfering RNA (siRNA) that specifically blocks this particular host cell protein resulted in a nearly tenfold decrease in production of West Nile virus (Fig. 2), as well as other flaviviruses.
revealed substantial transcriptional changes during Microarray analysis by our laboratory and by others cellular events during transformation because oncogenes are often transcribed in vitro upon infection of primary or life-extended endothelial cells with KSHV [27]. This in vitro model represents an excellent system to study cellular events during transformation because oncogenesis can be reproducibly triggered by infection with KSHV. Microarray analysis by our laboratory and by others revealed substantial transcriptional changes during spindle cell formation upon infection with KSHV [28,29]. Our study identified upregulation of the oncogene c-Kit. This host gene is particularly interesting because it is involved in other non-viral cancers, some of which also form spindle cells (e.g. gastrointestinal stromal tumors). Importantly, signal transduction by c-Kit can be inhibited by the receptor, tyrosine kinase inhibitor imatinib mesylate (STI571; Gleevec) [30]. Indeed, STI571 treatment prevented KSHV-induced spindle cell formation in vitro [28]. This observation suggested that the anti-tumor drug Gleevec might be useful in treating KS patients. This prediction was borne out in a recent clinical trial (http://www.asco.org/ac/1.1008_12-002489-00_18-00203-00_197-00103219-00_29-00A,00.asp).

Treatment with Gleevec of cutaneous KS in AIDS patients resulted in clinical and histological regression of lesions. These results suggest that c-Kit [and platelet-derived growth factor (PDGF) receptor; another target for Gleevec that is also present in KS lesions] represents a valid target for KS therapy.

Gene knockdown approaches to virogenomics
The examples listed earlier demonstrate that, in principle, functional genomics methods can identify new antiviral targets or targets for the treatment of diseases caused by viruses. However, finding the ‘needle’ in the ‘haystack’ of changes that occur in host cells during virus infection can be a daunting task. Until now, the availability of specific inhibitors was one of the reasons for choosing to study a particular host cell gene that was upregulated in a virally infected cell. However, recent years have seen a dramatic improvement in gene knockdown methods using antisense RNA or RNA interference (RNAi) [31,32]. Compared with earlier generations of phosphorothioate antisense molecules, which displayed many nonspecific effects, new antisense oligomers are highly specific and long-lived. For instance, phosphorodiamidate morpholino oligomers are highly resistant to enzymatic degradation and can be used to block translation of mRNA [33]. The most recent method for analyzing gene inhibition studies has been RNAi or siRNA, where small 21–23-nucleotide (nt) RNA duplexes interfere with the transcription program by directing degradation of homologous mRNA [34–37]. RNA gene silencing mechanisms are common to plants, fungi, nematodes, Drosophila and mammalian cells [34–36]. The mechanism by which siRNA cleaves the target sequence has not been completely defined. The current model proposes that dsRNA is cleaved into 21–23 nt in an ATP-dependent process by the RNAse III enzyme Dicer [34,36–38,39]. The 21–23-nt siRNA are thought to form a enzyme complex – the RNA induced silencing complex (RISC) – that unwinds the duplex siRNA and targets the complex to the recognition site on targeted mRNA where cleavage occurs and mRNA is degraded [37,38,40,41]. In mammalian cells, synthesized 21-nt siRNAs can lead to sequence-specific mRNA degradation without inducing the antiviral interferon response activated by longer dsRNA molecules [35].

Both new-generation antisense oligomers and siRNA can be used to knockdown expression of genes that were found by DNA microarrays to be upregulated in virally...
infected cells or organisms. What makes the combination of microarray and knockdown methods particularly attractive for viral systems is that, instead of *de novo* induction, microarrays usually indicate several-fold changes in the level of a given transcript upon infection. Thus, it is not necessary to completely shutdown expression of a target, but it is sufficient to reduce the level of induction back to the pre-infection state. Using the *in vitro* KS model, we applied morpholino antisense DNA oligomers against c-Kit to KSHV-infected endothelial cells. A single treatment was sufficient to inhibit foci formation, a process that takes several days to develop [28]. Similar results were obtained with siRNA against c-Kit (Fig. 3). Having established that antisense approaches work in this system, we are now using these methods to evaluate a large number of host genes found to be induced by KSHV on transformation of endothelial cells. At present, we have examined $\geq 30$ different host genes induced by KSHV. In several instances, we observed an inhibition of the transformed phenotype comparable to the c-Kit antisense or siRNA (A. Moses and K. Früh, unpublished observations). The corresponding genes represent known or novel oncogenes and potential new targets for therapy of KS as well as other cancers.

In other viral systems, siRNA targeted against viral or host cell genes has been successfully used to inhibit viral replication. The P gene of the non-segmented negative-strand RNA virus, respiratory syncytial virus, was selectively silenced and resulted in a decrease in viral progeny and loss of syncytia formation [42]. Poliovirus, a dsRNA virus, had reduced titers when treated with siRNA against capsid or viral polymerase genes, confirming that cytoplasmic RNA virus can be inhibited by siRNA [43]. However, poliovirus with a single nucleotide difference in the siRNA target region can lead to escape mutants, suggesting that multiple target regions need to be generated for therapeutic purposes [43]. Similarly, siRNA duplexes targeted against various HIV genes inhibited HIV replication [44–47]. In addition, decreasing CD4 levels by siRNA results in decreased HIV entry and spread [44]. Moreover, targeting of the downstream nuclear factor $\kappa B$ (NF-$\kappa B$) p65 gene, thought to be involved in stimulating HIV gene expression, decreased HIV-1 replication [47]. Similarly, human papillomavirus and hepatitis C virus replication were recently modulated with siRNA [48,49].

Different vector-based siRNA approaches have also been developed based on polymerase III promoter expression of RNAi molecules with small hairpin RNA structures that can be processed by Dicer [50]. This type of expression cassette has been incorporated into retroviral vectors for stable delivery and expression [50,51]. Recently, the therapeutic potential of siRNA was demonstrated by retroviral delivery of siRNA that could inhibit mutant K-RAS$^{v12}$ in human pancreatic carcinoma, leading to loss of tumorigenicity [51]. The therapeutic potential of RNAi *in vivo* has also been demonstrated in mice [52].

Although these results demonstrate that siRNA can inhibit a wide variety of viruses, it is our opinion that the real potential for this technique will be its use in validating which host cell products are important for virus replication or viral pathogenesis. As described earlier, pre-screening with DNA microarrays can identify potential host cell targets. However, an even bolder approach is to bypass microarrays altogether by using a whole-genome knockdown approach in a given system and examining subsequent virus replication ability. Although this approach is still difficult to implement in mammalian systems, owing to the large (and still unknown) number of genes and associated costs, a proof-of-principle study was successfully undertaken in *Drosophila* cells. Taking advantage of the completion of the *Drosophila* genome and that RNAi is extremely efficient in more primitive eukaryotes, N. Perrimon’s laboratory (Harvard Medical School) used a high-throughput RNAi screen based on 21 000 dsRNAs corresponding to every known *Drosophila* gene to monitor viral replication in Schneider cells. This approach revealed dozens of host cell gene products that were essential for viral replication but not for replication of the host cell (S. Cherry and N. Perrimon, pers. commun.). Once whole human-genome siRNA libraries are available, it will be possible to use similar approaches to identify the complete set of human genes required for viruses to complete their replication.

**Conclusions and future prospects**

Although most functional genomics studies of virology focus on cataloging global transcriptional and translational events, proof-of-principle studies are emerging demonstrating that some of the detected changes are events required or desired for virus replication or pathogenesis. Until now, validation could only be achieved using small-molecule inhibitors or dominant-negative proteins. However, the recent revolution in gene inhibition using small-molecule inhibitors will significantly facilitate the validation of a large number of potential targets identified by DNA microarrays or other methods. Moreover, whole-genome screens with siRNA will enable a complete evaluation and validation of the role of host cell gene products for viral replication. These approaches thus represent an unprecedented opportunity to characterize the functional host gene products essential for virus replication and disease.

**Acknowledgements**

We thank Alec Hirsch, Jay Nelson, Sara Cherry and Norbert Perrimon for unpublished information.
References