

Critical Dynamics in Host–Pathogen Systems

Arndt G. Benecke

Abstract Host–pathogen interactions provide a fascinating example of two or more active genomes directly exerting mutual influence upon each other. These encounters can lead to multiple outcomes from symbiotic homeostasis to mutual annihilation, undergo multiple cycles of latency and lysogeny, and lead to coevolution of the interacting genomes. Such systems pose numerous challenges but also some advantages to modeling, especially in terms of functional, mathematical genome representations. The main challenges for the modeling process start with the conceptual definition of a genome for instance in the case of host-integrated viral genomes. Furthermore, hardly understood influences of the activity of either genome on the other(s) via direct and indirect mechanisms amplify the needs for a coherent description of genome activity. Finally, genetic and local environmental heterogeneities in both the host’s cellular and the pathogen populations need to be considered in multiscale modeling efforts. We will review here two prominent examples of host–pathogen interactions at the genome level, discuss the current modeling efforts and their shortcomings, and explore novel ideas of representing active genomes which promise being particularly adapted to dealing with the modeling challenges posed by host–pathogen interactions.

Contents

1	A Systems Biology Challenge: Multiscale Integration.....	236
2	SIV Infection in Natural Hosts	238
2.1	Control of Chronic Immune Activation in Natural Hosts	240

A. G. Benecke (✉)

Centre National de la Recherche Scientifique, Institut des Hautes Études Scientifiques,
35 route de Chartres, 91440 Bures sur Yvette, France
e-mail: arndt@ihes.fr

2.2	Kinetic Proofreading as a Possible Mechanism for IA Control	241
2.3	The Importance of Timing Across Scales	244
3	Network Dynamics in Respiratory Virus Infections	245
3.1	Meta-analysis of Mouse Transcriptome Responses to Respiratory Viruses	245
3.2	Dynamic Interpretation of Gene Expression and Pathogenicity Correlation	247
4	Integration Over Time-Scales Using Probability Landscapes Over Genome Sequences	247
4.1	Requirements for a Mathematical Structure for the Object Genome	248
4.2	An Emerging Proposition for a Mathematical Genome Structure	249
4.3	Probability Landscapes	250
5	Concluding Remarks	256
	References	256

1 A Systems Biology Challenge: Multiscale Integration

After having generated high hopes and even more massive parallel data, *systems biology* is clearly on the verge of entering into a new phase to fulfill on the initial promise of revolutionizing not only the way we do biology but also our understanding of biologic phenomena (Tisoncik and Katze 2010). Success of this new phase will depend on solving some fundamental problems which so far have not, or only superficially been addressed, and will require more than ever a concerted and integrated effort spanning the entire spectrum of exact sciences.

The central problem we need to address is the integration of data and insights over multiple scales as to be able to make meaningful predictions about how complex traits and phenotypes emerge from assemblies of objects and the molecular mechanisms linking these objects on the one hand, and on the other, to be able to decompose phenotypes rapidly to understand the defining dynamics and their molecular basis. The former, inference-based analysis thereby actually encompasses also evolutionary questions, as most of the biologic systems we try to understand and describe are remarkably robust despite stochasticity being present, if not integral part of the mechanisms at multiple levels. The latter challenge of decomposition is still the main bottleneck on the road of designing therapeutical and vaccination strategies in biomedical research.

Decomposition and inference across time and space scales define the ultimate paradigm of systems biology research in as much as, if achieved and abstracted, the combination of both would lead to meaningful mapping functions from the object space to the phenotype space (Φ) and back (f) (Fig. 1).

The problem of integration over multiple scales is not unique to biology but also a major issue in physics and chemistry or social and economic sciences (Lesne and Lagües 2012). The problem, however, is particularly hard in biology, as the

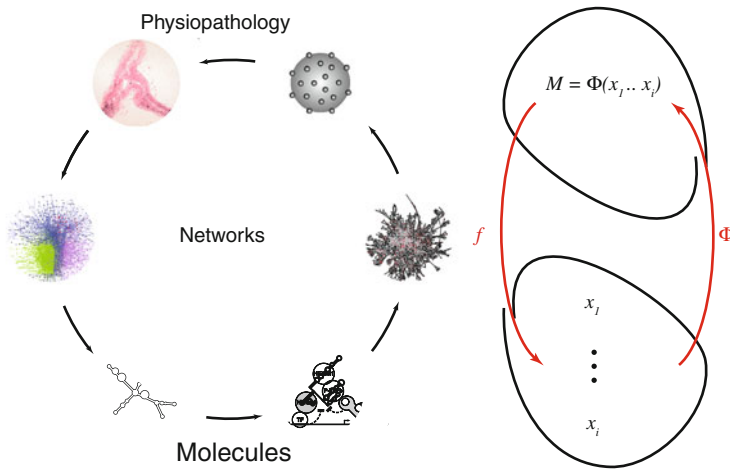


Fig. 1 Systems Biology Life Cycle: Decomposition of complex traits and phenotypes to understand the systems dynamics and the defining molecular objects and the mechanisms by which they interact; inference to make meaningful predictions as to how different objects interact to give rise to phenotypes and traits. Both processes will heavily rely on the identification and analysis of different biologic networks at different scales. The integration of information, objects, and their dynamics across scales represents the main challenge of systems biology today. Successful integration is the *sine qua non* requirement to identify and formulate the mapping functions ϕ and f from object space to phenotype space and back. Having a full set of these transforming functions would elevate the need to measure all objects and describe all possible phenotypes, and thus represent understanding of the system

integration has to be bi- rather than unidirectional. Consider a dune, thus a physical object—the dune’s physical properties depend entirely on the physical properties of the sand-corn. Using renormalization techniques, it is possible to mathematically describe a dune and investigate its properties under changing conditions (wind, humidity), without considering each sand-corn individually with simple equations such as the original Bagnold formula (Bagnold 1936). In biology, the physical properties of the molecular assembly such as a chromatin fiber will not only depend on the physical properties of the histones and the DNA, but in addition the histones and thus their the physicochemical properties have evolved under selective pressure acting on the chromatin fiber and its function (Benecke 2003, 2006; Bécavin et al. 2010). This symmetry established by the retrograde action of evolution is something which currently can not be captured by techniques such as renormalization (Lesne 1998, 2011), but will need to be accounted for in multiscale integration efforts. We have defined the term *function-dependent self-scaling* for models which describe for instance chromatin structure as a function of activity at the scale relevant to this activity (Lavelle and Benecke 2006).

Multiscale integration in biology is a fundamental problem for which currently little ideas exist how it could be solved. There are a few other problems of similar fundamental nature such as the role of stochasticity in biologic mechanisms and

how robustness of these mechanisms across changing environmental and systems-internal conditions can be maintained (Kaern et al. 2005). Interestingly, stochasticity here might be a solution more than a problem in many respects, but again a formal framework to describe, quantify, and predict such mechanisms is lacking. In what will follow, we will discuss some recent insights into functional genome representations to add a novel layer of investigation to the problem of gene expression regulation, chromatin structural dynamics, and genome structure–function relationships. These representations are thought to be particularly useful to compare genomes from closely related species and more importantly to provide new ideas of how to treat the case of two or more genomes operating together in a single cell such as is the case in infectious settings (Aderem et al. 2011; Tisoncik et al. 2009). To this end, we will first discuss two recent examples of successful network structure inference and dynamics analysis in systems virology, analyze the implications these results have for our thinking of genome function, and finally provide some ideas how to further investigate these systems using functional genome representations as a first step for a multiscale modeling effort.

2 SIV Infection in Natural Hosts

The definition of an effective HIV vaccine has only made modest progress despite prodigious efforts, as HIV successfully evades efficient and durable recognition by the human immune system (Ross et al. 2010; Belisle et al. 2011). Similarly, AIDS resistance in SIV natural host primates has been formerly believed to be caused by a lack of innate and adaptive immune recognition. This view is currently changing as four independent systems biology driven efforts have investigated in a comparative manner, the transcriptome dynamics in PBMCs and CD4+ cells of natural hosts for SIV as compared to Asian/New World primates that develop AIDS following SIV infection. Indeed, natural hosts just as AIDS progressor species display a rapid and strong innate immune response to SIV infection, and display all signs of successful immune activation (IA). The changes in the gene expression profiles are not only remarkably concordant between different natural hosts such as African Green Monkeys (*Chlorocebus sabaeus*) and Sooty Mangabeys (*Cercopithecus atys*), but also comparable in composition and strength to Rhesus Macaques (*Macaca mulatta*) and Pigtail macaques (*Macaca nemestrina*), the latter two being both AIDS progressors (Jacquelin et al. 2009; Bosinger et al. 2009; Favre et al. 2009; Lederer et al. 2009; Rotger et al. 2011). By systematic comparison of the gene networks indicative of IA between AIDS progressors and non-progressors not only common themes were identified, but also remarkable differences as to the duration of the innate immune response to SIV have been observed (Fig. 2). Indeed, IA in natural hosts ceases after the acute infection stage, typically after 2–4 weeks, whereas the gene networks driving the IA in AIDS progressors are still

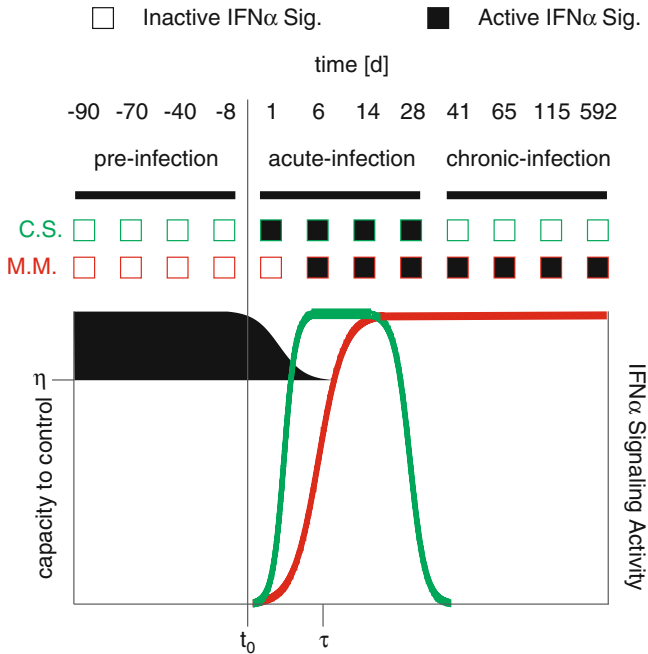


Fig. 2 Immune Activation in a Natural Host versus an AIDS Progressor—the West Coast Model. PBMCs from six African Green Monkeys (SIV Natural Host, *Chlorocebus sabaues*, here: “C.S.”) and six Rhesus Macaques (AIDS Progressor, *Macaca mulatta*, here: “M.M.”) were analyzed pre- and post-SIV infection at the indicated time points using transcriptome profiling and the activity of the Interferon α signaling pathway was inferred using ontology enrichment analysis (\square = predicted inactive, \blacksquare = predicted active, both at $p < 10E-3$) (Jacquelin et al. 2009). Two significant differences are observable: (i) C.S. control IA during the chronic phase of infection as opposed to M.M., (ii) C.S. seems more rapid in activating innate immunity than M.M. (Jacquelin et al. 2009). Similar differences are found in CD4+ cells from lymph nodes (Jacquelin et al. 2009), as well as other, independent studies involving a similar collection of data and different combinations of natural hosts and AIDS progressors (Bosinger et al. 2009; Favre et al. 2009; Lederer et al. 2009; Rotger et al. 2011). The recently proposed *West Coast Model* (Benecke et al. 2012) postulates that control of IA in natural hosts is a function of a mechanism reminiscent of kinetic proofreading (Hopfield 1974). Thereby, the capacity to control IA requires IA to cross threshold η before time τ . In the case of AIDS progressors, η is only reached after time τ , and thus the attenuation signal is not generated (a surfer missing the right moment to get on the board)

found active after the acute phase, and remain so until onset of symptoms of immunodeficiency (Bosinger et al. 2011, 2012; Manches and Bhardwaj 2009; Mir et al. 2011; Brenchley et al. 2010; Harris et al. 2010). Thus, it is the control of chronic IA, rather than absence thereof, which protects natural hosts from developing AIDS.

2.1 Control of Chronic Immune Activation in Natural Hosts

How can control, or absence of control in progressors, respectively, be thought to occur? Different hypotheses have been put forward, some of which can be disregarded or are unlikely to provide conclusive answers. SIV natural hosts do not display significantly altered infection or viral amplification rates and viral set-point titers. Moreover, chronically infected natural hosts maintain comparably high viral titers and can propagate virus. Viral particles isolated from natural hosts can be used to infect other animals (Jacquelin et al. 2009; Bosinger et al. 2009; Favre et al. 2009; Lederer et al. 2009; Estes et al. 2008). Thus, control of IA is neither directly connected to viral load nor is viral pathogenicity significantly altered during the course of infection.

The current hypothesis of how IA is attenuated in natural hosts is the presence of active signaling cascades which, upon a yet unidentified signal either attenuate IA in natural hosts or keep IA active in AIDS progressors. A logic table summarizes the four possible hypotheses depending on whether activators or repressors of attenuation or activation are considered (Table 1) (Bosinger et al. 2011; Harris et al. 2010). Currently, a specific search is underway in the different time resolved transcriptome profiles to identify such activators or repressors of either immune attenuation or IA, and which are differentially expressed/regulated in progressors and non-progressors. It will be of general, beyond the HIV field, interest to identify and characterize such activators and repressors which can promote or control chronic IA with obvious impacts for organ transplantation and autoimmune disorders (Rotger et al. 2011; Bosinger et al. 2011; Harris et al. 2010; Ye and Maniatis 2011; Lepelley et al. 2011).

The current generally accepted ideas on the control of IA in natural hosts, thus, postulate a necessary regulatory event (whether positive or negative) specific to either progressors or non-progressors. Thus, a dedicated signaling cascade composed of at least a sensor for a specific attenuation/activation signal, a transcriptional regulator, and a relay unit linking the sensor to the effector. Not only the molecules that are required specifically in either class of species, but also the nature of the specific signal pose a challenge in terms of evolution as an entire signaling pathway is required. Recall also that the signal for instance does not likely originate from the virus. Facing these dilemmas, we have recently formulated an alternate hypothesis for the absence of chronic IA in natural hosts which is based on a dynamic interpretation of the earliest innate sensing events following viral infection (Benecke et al. 2012). For the time being, this hypothesis is only modestly carried by direct experimental observations, as the time resolution with which early signaling events are usually studied is at least an order of magnitude above what would be required to directly assess the merits of the proposition. On the other hand, if this hypothesis, which appeals through its simplicity, would turn out to lead to the identification of a novel mechanism controlling long-term IA through early events, it would also define novel possible avenues for HIV vaccine development.

Table 1 Logic table for current hypotheses regarding control of IA in natural hosts

	Immune attenuation		Immune activation	
	N.H.	A.P.	N.H.	A.P.
Activator	+	–	–	+
Repressor	–	+	+	–

N.H. Natural Host, A.P. AIDS Progressor, + present, – absent

2.2 Kinetic Proofreading as a Possible Mechanism for IA Control

Kinetic proofreading is a potent mechanism known in molecular discrimination (Hopfield 1974). Kinetic proofreading is a process in which, through expenditure of additional energy, ligand recognition is split into two or more individual events in order to increase specificity and discriminatory capacity between closely related ligands or interaction partners with modestly different free energies of binding. In a first step, usually coupled to a conformational change in the receptor achieved through the hydrolysis of ATP, a candidate ligand is bound and presented to an independent interaction surface. Only if this second, independent interaction occurs rapidly enough, the recognition is conclusive, otherwise the ligand is released as the receptor snaps back into its original conformation. This mechanism has been studied in great detail theoretically and shown to drastically increase recognition of a *bonafide* ligand over analog molecules with very similar free energies. The error thereby is reduced beyond the thermodynamic bound—sometimes referred to as the specificity paradox upon which Hopfield based his predictions that ribosomes match codons and amino-acid-loaded tRNA anticodons using a kinetic proofreading mechanism. This has later been proved experimentally also for the way that aminoacyl tRNA synthetase operates (Hopfield 1974; Hopfield et al. 1976). Furthermore, and more relevant to this discussion, T-cell receptors use kinetic proofreading to enhance discrimination of *bonafide* ligands from closely related molecules to ensure correct signaling (McKeithan 1995). Finally, some evidence suggests that kinetic proofreading could also be found at the basis of RIG-I or TLR mediated recognition of foreign in innate immunity (Loo and Gale 2011; Liu and Gale 2010; Suthar et al. 2010).

For the sake of argument, let us assume that a strong and immediate innate immune response is not only a first line of defense to gain the required time for setting the stage for adaptive immunity, but that it is also a mechanism to proofread the adaptive immune response. In this scenario, some of the mechanisms of innate immunity would be required to be activated in order to maintain sustained, general IA beyond acute infection. Absence of innate proofreading would then lead to total inactivation of immune function. However, also the exact opposite effect might be at work—innate proofreading is required to attenuate continued IA. We believe that this latter scenario is more likely, and better reflects the general observations made about immunity. A typical pathogen will trigger

(many) different innate sensors simultaneously. The multitude of signals acts synergistically to mount the immediate innate IA which in turns triggers adaptive immunity. Maintaining this early response over prolonged periods of time, as observed in AIDS progressors, does not add any advantage to the system, however, is costly in terms of energy expenditure and precludes specific activation of downstream processes. If one of the different innate sensing mechanisms serves as proofreading mechanism, it makes more sense to propose that the proofreading is meant to attenuate the early innate response rather than sustaining or driving it as the latter would be redundant with the other mechanisms. In other words, the proofreading would simply signal that innate IA has been successfully triggered and thus needs to be attenuated in the near future in order to set the stage for adaptive immunity, avoid exhaustion of resources, and redundant signaling without added benefit.

Therefore, an innate sensing mechanism that triggers attenuation of IA would represent a simple feedforward control which does not require any additional specific signaling pathways or additional signals in order to be functional (see Goodman et al. 2011) for an interesting example of a feedforward mechanism in viral replication). This appears to be one strong argument in favor of the existence for such a dual purpose innate sensing that acts in one of those two aspects reminiscent of kinetic proofreading.

The second interesting argument can be formulated in favor of this hypothesis which is the dynamics of proofreading. As discussed above, through the addition of irreversible (energy consuming) steps prior to and integral part of faithful recognition a delay function is implemented. In other words, every one of the independent irreversible prerecognition steps needs time to complete; and thus the increase in specificity of recognition is not only 'bought' through energy consumption but also accompanied with varying delays between the initial encounter and positive recognition, which are a function of the number of successive prerecognition steps and physical proximity. In this context, the time delay creates a lag-time for the attenuation signal of innate activation which would prevent early shutdown. In other words, not relying on a specific signaling pathway for attenuation creates the problem that innate IA and its attenuation are triggered at the same time leading to conflicting signals. If, however, the attenuation signal is lagging behind because of its increased specificity, a functional feedforward repression is implemented (Benecke et al. 2012).

Finally, the dynamics of such a proofreading mechanism could potentially also explain the differences observed between natural hosts and AIDS progressors following SIV infection. As a matter of fact, a kinetic proofreading mechanism defines two boundaries on time. First, discussed above, there is a lower bound for the recognition process defined by the delay in time over the one or several irreversible steps. But also a second, upper bound, on time is explicitly part of the mechanism. If the recognition step n is too slow compared to the step $n - 1$ the process aborts as unsuccessful. Hence, the execution time for step n is bounded by a function of the off-rate of $n - 1$. Practically speaking, the hypothesis presented here suggests that there exists a window of time during which recognition has to

occur in order to trigger attenuation of innate IA. This window of time starts with the earliest prerecognition event at t_0 (infection) and continues up to some upper limit τ which has to be sufficiently close that robust (a significant fraction of a large number of events) recognition can occur. If this recognition occurs too late, the attenuation signal can no longer be released and IA continues chronically. This is a strong hypothesis which should be verifiable experimentally. Indeed, there even seems evidence in the existing transcriptome profiles for early dynamics playing a key role in the attenuation of IA in natural hosts, and why immune attenuation does not occur in AIDS progressors (Fig. 2). Indeed, it appears that innate IA occurs more rapidly in the natural host AGM as compared to Rhesus Macaques if the ontology-based inference of the activity of the interferon α pathway is accepted as a proxy (Jacquelin et al. 2009). The lower schematic illustrated the two main differences in the activation and attenuation kinetics between the AGMs (green) and the Macaques (red) and also schematizes the window of opportunity (black) for a feedforward attenuation mechanism reminiscent of kinetic proofreading. The threshold η needs to be crossed by the early recognition events before τ expires (see above) and too slow IA in the case of AIDS progressors (red), albeit sufficient in amplitude to cross η , fails to do so within the window of opportunity set by the proofreading mechanisms' upper and lower bounds on time. Note that, we assume here that the lower bound is defined by the first encounter with viral particles/components thereof or immediately after, thus is identical for the two species in this experiment, and that the upper bound is a function of the intrinsic lifetime of prerecognition complexes assumed to be identical in both cases as well. Thus, the only variable in the system is the speed with which IA occurs in both species. This can be viewed as analog to the situation of a surfer. If pathogen encounter and innate recognition as foreign is considered a wave at the beach, then IA could be seen as a surfer getting up on his surfboard. If the surfer fails to mount during the window of opportunity (defined by the width of the wave-back, thus intrinsic to the wave), the surfer will sink; thus, the term west coast model used (see Benecke et al. 2012 for a detailed discussion on this argument). Relevance of this model stems from the following observations: SIV-infected NHPs and HIV-infected human AIDS progressors mount their innate immune response too slow or rather too late leading to a non-attenuation and thus chronic IA. This unresolved innate IA wears down the system and leads consequently to decline in CD4+ T-cells, the hallmark of AIDS (Pandrea et al. 2011). Natural hosts for SIV on the other hand, such as sooty mangabeys, African green monkeys, and mandrills display timely responses to infection leading to successful IA and concomitant IA attenuation and, due to absence of specific humoral responses long-term tolerance of the virus (Jacquelin et al. 2009; Bosinger et al. 2009; Favre et al. 2009; Lederer et al. 2009; Rotger et al. 2011).

Comparative transcriptome profiling between an SIV infected natural host (here: *C. sabaueus*) and a progressor (here: *M. mulatta*) shows evidence of a lag-time of IFN α (as proxy for innate IA) signaling in progressors (Jacquelin et al. 2009) (Fig. 2). Note that, this delay of about a week might, however, be due to phenomena not necessarily related to the kinetics of IA, as the amplification

kinetics of the two adapted SIV viruses might be different, or for instance, we do not know whether or not the effective doses might be different between the two species. Still, it seems unlikely that such before mentioned effects would entail such profound changes in the IA kinetics, and thus this experimental finding might be regarded as a potential support of the proposition of kinetic autoattenuation of IA in natural hosts. It will be of outmost interest to better characterize the activation dynamics across the entire spectrum of known natural hosts and progressors in order to contrast possible differences in the activation kinetics with human subjects (or more likely *ex vivo* cellular models) representing the different observed classes [progressor, long-term non-progressor (LTNP), elite controller (EC)] as especially the LTNPs would be candidates of having acquired a similar attenuation mechanism as natural hosts. In this context, particular attention should also be given to the investigation of co-infection schemes with different pathogens (Schreiber et al. 2011). This would then also lead to the proposition that, similarly as to non-human primates, it is not the absence of an effective adaptive immune response to HIV itself but the failure to control the innate immune response which is the main driver of AIDS.

2.3 *The Importance of Timing Across Scales*

In conclusion, the proposition of mechanisms similar to kinetic proofreading for the coupling between innate and adaptive immunity is appealing as it combines simplicity with fidelity. Thereby, innate IA, with its obvious role of identifying foreign from self, would in the same time serve as a guard against inappropriate initiation of adaptive immunity by automatically attenuating the primary response. In order for this model to work, however, one needs to evoke the concept of a fading capacity to attenuate IA, and postulate that the attenuation threshold η is never reached in AIDS progressors in time τ (Fig. 2). Conclusive insights on the model presented for the coupling between innate and adaptive immunity, and the propositions regarding SIV and possibly HIV infection will require the successful translation of molecular profiles such as the transcriptome profiles obtained in the four cited studies into a dynamic view of the host's cellular immunity. This might sound simpler than it indeed is for several reasons such as experimental limitations imposed by the model systems or the technologies at hand for monitoring molecular events and their proxies (mRNA, signal cascade activation, metabolic activity), but mainly as one will need to overcome the problem of integration over multiple scales from the dynamics of single molecular events (in the micro- to millisecond range) to events at the organ level occurring on the scale of hours to days (please refer to the remarks made in Sect. 1). After having discussed briefly the second example of the importance of the network dynamics in immune responses from respiratory virus infections in Sect. 3, we will develop some ideas of how this general problem might be partially solvable for the particular cases discussed here (Sect. 4).

3 Network Dynamics in Respiratory Virus Infections

Other chapters of this volume discuss in great detail the case of different respiratory viruses and their interactions with their native hosts. We will, therefore, discuss here only a single finding from recent work on a meta-analysis of host transcriptome responses to a compendium of essentially Flu and SARS infection scenarios. As will be seen below, the observation made by Chang et al. (2012) pertains to host response dynamics, similarly as the studies discussed with respect to SIV and the innate IA in different hosts. Distinctively, the respiratory virus example does not compare different hosts for the same of differently adapted viruses, but rather different viruses (or pathogenic states) in a single host.

3.1 *Meta-analysis of Mouse Transcriptome Responses to Respiratory Viruses*

The threat of a highly lethal viral pandemic remains a major threat; the recent SARS-CoV 2003 and the H5N1 pandemics testify to the uncontrolled potential of emergence of respiratory viruses with possibly devastating characteristics reminiscent of the 1918 Spanish Flu (Donnelly et al. 2003; Beigel et al. 2005). Accordingly, major efforts are directed toward an understanding of the viral determinants of pathogenicity and their possible horizontal drift on the one hand and possible restriction factors or key modulators of pathogenicity on the side of the host on the other.

Deriving robust and unique molecular fingerprints for physiopathologic phenotypes from massive parallel experimental data is not only of extraordinary value for the understanding of pathogenicity but also a serious challenge given the current absence of systematic procedures (Ein-Dor et al. 2005). Biologic variability and insufficient sampling of the relevant state-space at present preclude formal approaches to molecular signature definition. A molecular signature is best defined using the isolation principle (Gregorius 2006) as the minimum number of biologic observables required to (i) discriminate the studied phenotype from some (ideally: any) other existing phenotype (external isolation), (ii) differentiate sufficiently between replicate analyses of the same phenotype thereby capturing biologic variation (internal cohesiveness), (iii) be robust against technical and biologic variability, and (iv) be of biologic relevance by representing the underlying more complex phenotype in its principal characteristics.

In order to advance in the definition of the hallmarks of lethal infection by respiratory viruses, Chang et al. compiled a compendium of published individual transcriptome studies on mouse lungs in order to identify gene signatures which obey by the definitions set forth above. The compendium of microarray data from the 12 analyzed studies was composed of a total of 733 individual transcriptome profiles, roughly equally distributed over the three physiopathologic groups

(‘high’, ‘medium’, and ‘low’ pathogenicity) and their corresponding controls. Four different methods of meta-analysis stemming from two different philosophical approaches were used and compared in their absolute and relative performance. Processed data were either converted to logratios to identify genes that show opposite regulation in HPI and LPI, or directly submitted to meta-analysis by direct comparisons. In previous studies, both targeted and genome-wide approaches have been used to identify particular host pathways deregulated during infection. In parallel, a direct comparison of gene expression in ‘high’ and ‘low’ pathogenicity groups was performed. Statistically significantly differentially expressed genes were compiled to result in a characteristic gene signature when comparing the initial ‘high’ and ‘low’ groups. The fundamental difference between the three earlier, logratio based methods, and the latter direct comparison signature is the implicit choice of reference gene expression levels as well as the subsequent classifier used to choose signature genes. While the former methods will select for those genes that are uniquely/oppositely regulated in ‘high’ versus ‘low’ pathogenicity settings, the latter will select for genes that are statistically significantly differentially expressed between both conditions. The logratio meta-analysis derived signatures could be, in accordance with Sonnenschein et al. (2011), referred to as ‘digital’ and the direct comparison signature which comprises both gene IDs and gene expression values as ‘analog’. All of the pathogenicity signatures were then compared among each other and characterized individually toward the objective to characterize responses that were present across high-pathogenic infections (HPI) and low-pathogenic infections (LPI).

The analog pathogenicity signature (aPS), correctly classified test data from the comparison of infection with one of two swine-origin influenza virus A strains, pandemic H1N1 (CA/04), or a mouse-adapted lethal variant (MA1 CA/04) (Bradel-Tretheway et al. 2011) not comprised in the initial compendium used for the competitive meta-analyses. In-depth analysis of the aPS revealed, furthermore, that biologic conditions classified as intermediate between HPI and LPI often belonged in the case of MPI data to late time points after infection, and for HPI data to early time points, leading to an analog immune response model for respiratory virus infection. The aPS derived by comparative meta-analysis of this respiratory virus infection compendium can be, thus, used to correctly classify host transcriptome responses according to clinical pathogenicity. The reason why the aPS outperforms the alternate digital pathogenicity signatures derived through the other three meta-analysis methods is explained by the striking observation of an analog that is continuous and correlated, host gene expression response to pathogenicity. Gene expression of this continuous response can be either positively or negatively correlated with pathogenicity, the latter being only recently recognized to exist (Kash et al. 1918; Cilloniz et al. 2010). This finding has not only technical implications for molecular signature definition strategies, but also for the understanding of the physiopathology of respiratory virus infection: continuous responses of gene networks to pathogenicity rather than different or oppositely regulated networks specific to ‘high’ or ‘low’ pathogenicity dominate the immunologic response of the host to viral infection which has major implications

for medical targeting of these networks. On the other hand, the observation of analog immune responses lends hope to the successful identification and boosting of host innate and adaptive immune mechanisms against high pathogenicity infections.

3.2 Dynamic Interpretation of Gene Expression and Pathogenicity Correlation

Important in this context is the possibility that infectious outcome might be encoded by the activation dynamics of host response gene regulation. In other words, one might have a hard time to find genes specifically responding to HPI or LPI, but rather only different activation dynamics for genes regulated in either case. Figure 3 illustrates the possible underlying mechanisms for such an observation.

Comparative meta-analysis of the host transcriptome dynamics following infection with high- or low-pathogenic respiratory viruses identified a gene signature characteristic of the pathogenicity of the virus (Chang et al. 2012). Highly pathogenic viruses such as influenza A subtype H5N1, reconstructed 1918 influenza A virus, and SARS-CoV thus illicit the same immune reaction than low- and medium-pathogenic viruses, however, to a higher degree. The observed strong correlations with pathogenicity could originate from two different, dynamic regimes of the underlying network (Fig. 3).

In conclusion, the meta-analysis of transcriptome profiles from respiratory virus infections reveals again critical dynamics of innate immunity at time-scales below currently investigated scales. The possibility of similar mechanisms at work when comparing the case of SIV infection in natural hosts (Sect. 2) and respiratory virus infections in mice (Fig. 3 right), possibly even further strengthens the general idea of time dynamics being of critical importance to host–pathogen interactions. In the following section, we will ask how such dynamics can be better inferred and analyzed using novel genome representations.

4 Integration Over Time-Scales Using Probability Landscapes Over Genome Sequences

In what follows, we will discuss a recent proposition for a mathematical description of a genome and associated activities. We will first argue for the need of such a structure, then discuss the general outline of the recently proposed structure, and finally discuss how this structure might help to further the concepts

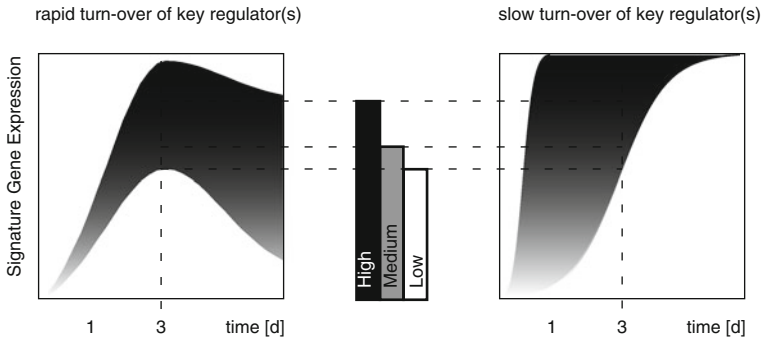


Fig. 3 Two alternate dynamic interpretations of the observed strong correlation between gene expression activity and pathogenicity (Chang et al. 2012). The uncovered positive and negative correlations between mRNA levels produced from a signature set of genes relevant to respiratory virus infection in mice with the corresponding pathogenicity of the virus (viruses or conditions were attributed to one of three discrete categories ‘high’, ‘medium’, and ‘low’, center) have two possible mechanistic origins. First, as initially proposed by Chang et al. (2012), while variable in time, a given gene at any given moment will be expressed as a function of viral pathogenicity (*left*). Second, it is also possible that all the signature genes will share similar expression values independent of the pathogenicity of the virus, in this case, however, at different moments in time (*right*). These regimes are not necessarily exclusive. Note that with the current resolution of the existing data a direct inference of which of the two regimes actually at work is impossible. Note also that the identification of which of the two mechanisms is at work would lead to strong, testable hypotheses, and provide directions for future experiments aiming at dissecting the gene regulatory network(s) relevant to the viral pathogenicity. The identification of the key regulator(s) driving the effective network and its dynamics were greatly facilitated if one could make a prediction as to the turnover of these regulators (which can be estimated from the time-series data for all genes). Note finally that the regime described on the right—disparity in activation (and symmetrically repression, not illustrated)—resembles the observations made in the case of comparative SIV infection in natural hosts versus AIDS progressors (Sect. 2, Fig. 2, opening the exciting possibility of a similar, if not identical, phenomenon taking place in both scenarios)

discussed in the two examples above (Sects. 2, 3) by providing a basis for the decomposition and inference over multiple time-scales (Sect. 1).

4.1 Requirements for a Mathematical Structure for the Object Genome

Today, genome biology is essentially based on (linear) statistical approaches. This is somewhat surprising as the amount of available information and experimental data is not, nor likely will ever be in the near future, sufficient to derive proper statistics on the object ‘genome’. The large number of different biologic conditions will not be exploitable and the space of biologic conditions hence will remain extremely sparsely sampled. Furthermore, it will almost nowhere reach sufficient

density (e.g. recordings of many independent biologic replicates) to allow proper statistics. Moreover, simultaneous observation of all relevant determinants at all relevant scales over time is not possible, the experimental data will remain independent observations. Statistics on those will not enable to construct causal links rather than correlations between them. Furthermore, standard statistics is inappropriate for the questions posed since biologic processes are not generic, and arguments of parsimony, typicality, and natural chance of occurrence fail. Finally, statistical descriptions *per se* do not provide causal relationships, and hence do not provide comprehension of the underlying mechanisms. There are no obvious computational remedies to these limitations due to the evolutionary (and possibly other) feedback from the level of the higher, emergent scale down to the molecular scale as discussed in [Sect. 1](#) (Moore 1990; Israeli and Goldenfeld 2006).

The object genome (which includes all of its possible activities) is likely to be ‘computationally irreducible’ (Moore 1990), meaning that if we aim at computing the behavior generated by genomic information, we have to perform as many operations as there are time steps, elements, and interactions. There is, hence, little possible reduction of the complexity of the biologic system *genome* by computational methods unless a unified, mathematical self-consistent structure can be formulated. *Time* will be one important but not necessarily privileged dimension of such a structure.

4.2 *An Emerging Proposition for a Mathematical Genome Structure*

In order to go beyond statistical approaches and, thus, to reach a level of understanding of genomes which is sufficient for meaningful inference of regulatory processes the current concept of a letter-based alphabet for genome coding needs to be revisited. Comprehension, or at least the possibility of inference of networks and their dynamics over multiple scales is likely a prerequisite to targeting multifactorial diseases such as cancer, genetic disorders, or pathogen-induced malignancies. The examples discussed above illustrate well the limitations of current methodologies at hand. Let us, thus, first recapitulate the main features which need to be captured by mathematical (or functional) genome representation: a genome (i) codes for a number of molecular machines that catalyze elementary biochemical reactions, and (ii) has evolved to orchestrate the molecular machines in a manner that whatever form the organism takes in response to external or internal stimuli the organism remains alive (Benecke 2006). This seemingly trivial concept that any transitions from one functional (active) form of the genome/organism to another can only happen at the condition that any intermediate represents a viable genome/organism needs to be exploited as it is the strongest constraint on the system. The true ‘miracle’ does not lie within the elementary machines but within

the fact that they self-organize across different time and space scales into a functional form whether it be at an embryonic or an adult state (Smet-Nocca et al. 2010). It is the rules of interaction (direct or indirect) that are at the essence of the genome. These rules of interaction are coded in the genome at its sequence level, but also on the level of its structural and spatial dynamics (for instance: activity-dependent subnuclear localization, or localization-dependent activity). Thereby, any elementary information in the genome (such as a single nucleotide) has a role (even seemingly negligible) of coding for any part of the functional forms of the genome at different time and space scales (Benecke 2006). The functional forms of a genome are thus expressed through nonzero contributions (weights) from individual elements which interact within a highly constrained, hence rigid structure. Note that from a computational viewpoint, an active genome is presumably a universal Turing machine (Benecke 2006). Recently, an initial proposition for a mathematical representation has been made where nucleotide frequencies as well as measurements on the activity of any part of the genome under defined biologic conditions are simultaneously expressed as probability distributions (Lesne and Benecke 2008a, b). This mathematical structure allows, which yet also has some questionable properties, see below, allows to introduce concepts from algebraic geometry for data analysis and modeling. We thereby use three independent paradigm shifts which lead to a modified approach to the inference problem in functional genomics (Benecke 2008).

4.3 Probability Landscapes

A genome is currently represented as a string composed of a four to six (DNA methylation, gaps) letter alphabet. Most approaches consist of identifying meaningful ‘words’ within this text, often by trying to identify over-represented subsequences that coincide with measurable quantities or changing quantities such as a gene, the amount of RNA transcribed from a gene, or the presence of a gene regulatory factor or particular chromatin modifications associated with the studied process in a given biologic condition. The genomic sequences obtained over the past decade reveal a low complexity of the genomic sequence, especially in non-coding regions, and consequently high-fidelity statistical inference of functional elements is essentially limited to protein coding sequences which account for only $\approx 2\%$ of the total human genomic sequence. Paradoxically, even what was considered to be a well-defined concept, the notion of a gene, is being challenged by the recent discovery of short and long, untranslated RNA sequences (microRNAs, ncRNAs), and the discovery of increasingly complex patterns of alternate promoter and splice-site usage. The concept of probability landscapes replaces the one-dimensional view of a genome by a stacked structure over genome positions, where the stack contains the representation of all biologic objects and events relative to the position n along the genome (Fig. 4). This mathematical structure gives at the same time the framework to

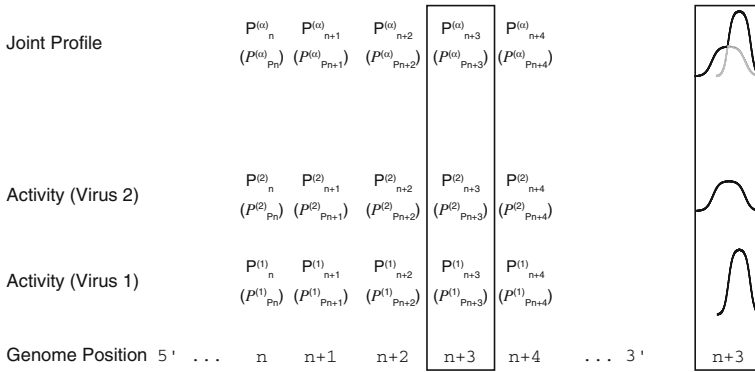


Fig. 4 Probability landscapes, which include as reference set the probabilistic representation of the genomic sequence obtained from several to many individuals, can be used to discover and analyze longitudinal correlations efficiently among the initially heterogeneous and unreliable descriptions and genome-wide measurements. The structure consists of probability density distributions stacked on any genome position n defining the vertical extension. Horizontally, along the one-dimensional genome, a layer is generated for every biologic condition and every experimental measure. In this schematic representation, the probability distributions for two measures of activity of two different viruses over a five base genome is illustrated (Lesne and Benecke 2008). These profiles than can be integrated vertically (schematized on the right) using appropriate formalisms. A large collection of such geometric and algebraic ways to generate what is here referred to as joint profile exist (Lesne and Benecke 2008)

analyze data, to reconstruct missing information using rigidity-like and coherence arguments, and to express inherently multiscale causal relationships that can be used to explain genome function. Mathematical does not mean abstract, since on the contrary any set of experimental data or concrete interactions are transformable into the probability distributions (Lesne and Benecke 2008a). In turn, the probability distributions used allow the inference of a more integrated knowledge without having to prescribe all local properties and connected relationships. Rather than considering individual states of an active genome, probabilities describe the relevance of any object mappable to the genome (for instance: physical properties of chromatin, or transcription factor binding) to these states (Lesne and Benecke 2008a). As any relevant information on all levels, features (objects such as genes, regulatory sequences), and experimental data can be expressed as probabilities, a unified representation is obtained. The *ensemble* of probability distributions at site n constitute the stack and horizontally, thus over all positions n_i , a profile. Finally, rather than focusing on objects and states (or their probabilities) the aim of this form of representation is to be able to access the transformations between the probability distributions that govern their mechanistic, biologic relations. The set of transformations thereby constitutes the mapping functions f and Φ from Fig. 1 for the phenotypes associated with genome activity provided sufficient data have been integrated.

Probability landscapes provide, thus, a unified structure consisting of probabilities $(P_n)_n$ and associated quality estimates $(P_{P_n})_n$ —in the form of functional probability densities (probabilities of probabilities)—to integrate any type of relevant genomic information into a coherent annotation. Most importantly, genomic sequence itself, its annotation with empirically derived features such as genes and regulatory sites, and any type of functional genomics data can be described in this manner. The rationale of this probabilistic description is not necessarily to account for an underlying stochasticity, though for some biologic processes this is indeed relevant, but rather to provide an efficient way to formulate partial knowledge and turn relative data of very heterogeneous nature and origin into absolute values and a homogeneous representation of the initial observations. Genome probability landscapes are systematic as any type of relevant information can be correctly and sensibly projected upon the genome positions. This projection has a single nucleotide resolution, producing a (at least locally) continuous profile. The proposed framework is coherent, as any information is converted without exception into the very same structure: probabilities with associated probability densities for local quality estimation. While the proposed representation of information is far from optimal in terms of compression, it provides a direct, systematic, and coherent interface for analysis, thus rendering numerical calculation efficient. The systematic nature of genome probability landscapes and their coherent structure allows easy exchange of information between different research teams. The simple structure of the resulting data also makes the framework easily portable between different computing environments as there is no real need for a solid database structure to generate, store, and handle the information provided that the same metrics are used to generate the profiles. Note that this aspect is a little oversimplification, as using the same metrics is not trivial when all aspects of quality control of the raw data, missing value imputations, and normalization have to be considered. It also appears that the concept is future compatible, as any type of relevant information can also be included in the very same manner into the existing landscapes (we disregard here whether or not this information makes previous data obsolete). This latter point is certainly of heightened interest giving the speed at which technology is developing for instance with respect to ‘deep’/next generation-sequencing (NGS) and digital PCR. A structure that thus can meaningfully combine ‘old’ e.g., microarray type of data with ‘new’ NGS data will reduce the requirement for rerunning the same biologic conditions with the latest technology. Finally, the proposition to use probability landscapes for the integration of such data is—as it is inspired by and organized along the DNA sequence—a natural solution. Importantly, probability profiles can also accommodate the description of physical properties of DNA (for instance bending and intrinsic curvature) and chromatin fiber (local elastic constants, compactness), as well as the conformation of its nucleosomes and topologic constraints (conserved linking number within a loop); all these features are expected to play a key role in for instance transcriptional regulation (Widom 1998; Lesne and Victor 2006; Lesne and Benecke 2008a). Even nuclear dynamics could possibly be expressed

through the location, either central or peripheral of chromatin loci within the nucleus (Spector 2003; Cabal et al. 2006).

4.3.1 P-Landscape Based Analysis

Genome probability landscapes essentially provide the first step into processing any raw experimental data into a unified expression suitable for systematic genome-wide integration and analysis. To reduce unnecessary formal, mathematical, and computational complexity, we have developed methods for collapsing subsets of the landscapes whose basic step is an analysis of the stacks at a given genome location n (Lesne and Benecke 2008b). In the toy example given in Fig. 4, one might for instance want to ask whether it is necessary to consider the activity profiles of Virus ~ 1 and Virus ~ 2 as distinct or whether it is more meaningful to pool them. In other words, does the profile of Virus ~ 1 when jointly considered with the one of Virus ~ 2 provide independent information which needs to be considered or can the one be used to rather back the other? To answer, a measure called Kullback–Leibler divergence (Kullback and Leibler 1951) can be employed to measure the relative contribution of either activity profile to the joint profile. Each individual profile's weight to the combined measure is obtained using the average presumed frequency of these subsets (rather subpopulations). This amounts to one example of a vertical comparison which can be performed along the genome. Then, a longitudinal integration of the local divergencies is performed along genome regions of relevance (e.g. over the location associated to a given gene) allowing to analyze the feature divergence profile of a biologic condition over the entire genome or defined intervals. This genome-wide distance measure is meaningful, unlike the individual feature profiles. If the conditioning by any combination of individual or averaged profiles leads to a statistically significant divergence (suggesting that the associated subpopulation is well delineated and has a specific signature) the profile is kept as a separate entity. In contrast, if statistical significance is not reached, the condition is considered non-pertinent to the biologic question posed as it does not provide a measurable constraint on the value of the joint profile and can be combined with any other statistically insignificant conditions. This process, thus, integrates and thereby collapses part of the landscape to restrict to statistically divergent information (whether this is also biologic meaningful information can not be determined at this stage). Two advantages arise in this case: (i) the complexity of the structure is reduced in a controlled manner in so far as it is irrelevant to the biologic question investigated, and (ii) the statistical power of the joint probability profile is increased. As shown in Lesne and Benecke (2008b), this procedure can be performed at any interesting scale or functional level and thus the probability landscape over the genomic sequence can be reduced in complexity until all remaining context-dependencies reach statistical significance at which an optimum for computational complexity and statistical power is reached. Different biologic conditions can thereby be defined with maximum

flexibility using separate or overlapping subsets of subconditions in a hierarchical manner. The Kullback–Leibler divergence-based method discussed in Lesne and Benecke (2008b) represents, thus, a systematic and simple way of testing the statistical limits of complexity reduction and hence explanatory power of the integrative genomics data in their respective contexts. Note that since we are comparing the distributions of the same random variable under different conditions, it is only the distance (or divergence) between the two distributions that is meaningful. A joint probability, such as mutual information, could not be envisioned. This also holds for the case of two different variables because the joint probability distribution is inaccessible. From a general perspective, our method represents an application of concepts related to context trees to the probability landscape idea. Context analysis and landscape collapse thereby operate in similar manners to Markov chains with variable length for the analysis of time-series and historic context (Bühlmann and Wyner 1999; Maubourguet et al. 2008). We also note that the Kullback–Leibler divergence calculation provides measures that can be used directly for clustering of probability profiles. Clustering of probability profiles might help to establish and analyze relatedness among data otherwise not compared directly.

4.3.2 Expressing Time in P-Landscapes

As discussed earlier (Sect. 1), the successful integration of time over scales is one of the current bottlenecks of a systems biology description aiming at a discovery mechanism for mapping functions between objects and phenotypes. The two cited examples from virology (Sects. 2, 3) underline the potentially crucial importance of molecular dynamics and their coupling to macroscopic behavior. There are two different possibilities to incorporate time into probability landscapes. First, explicit integration using which will be based on directly using the different time points from the kinetic, to stay within the perimeter of the examples from above, transcriptome profiles to generate individual probability profiles now depend on time: $P_n^{(\text{Virus 1})}(t)$ (probability to observe activity of Virus 1 in the experimental condition at site n and time t). It is then possible, generalizing the methodology developed for single time P-landscapes to compare those using for instance the Kullback–Leibler formalism, to align profiles from different biologic conditions (Virus ~ 1 vs. Virus ~ 2) using mutual information optimization to determine a local or global shift (compare Fig. 2), and finally fit a model of the evolution over time using a stochastic operator.

Alternatively, time might be captured only abstractly, and thus indirectly. Consider once more, the schematized behavior of the respiratory virus induced host response signature from Fig. 3. Whatever the interpretation of the experimentally measured result (center), thus whatever the underlying mechanism (rapid or slow turnover of key regulator) in both scenarios a density (here: pathogenicity) function over time is at the origin of the measured result. As discussed above,

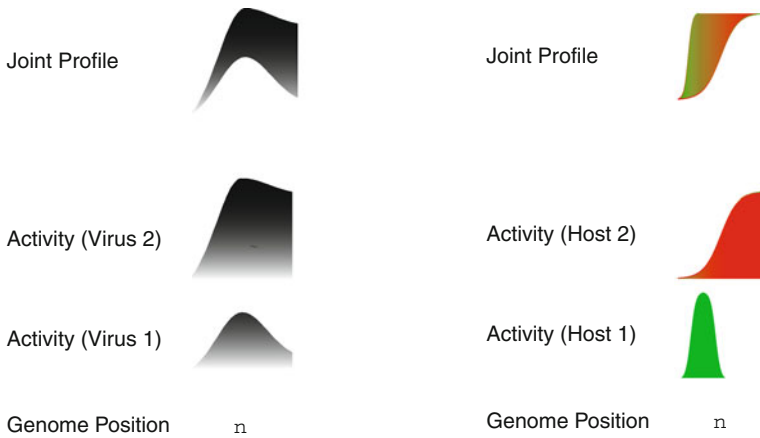


Fig. 5 Capturing time abstractly within the framework of probability landscapes (Lesne and Benecke 2008a, b). Both proposed mechanism (rapid or slow turnover of key regulator) which would lead to the remarkable correlation (and anti-correlation) between the expression levels of key signature genes for respiratory virus infection as a function of the pathogenicity of the analyzed virus lead to density distributions of gene activity with respect to time. These density distributions are characteristic for the virus and can be expressed as probability profiles along the host genome (here illustrated for a single genome position, which might be as discussed in Sect. 3, either indeed a single nucleotide or a consecutive stretch of the genome associated to a measured activity—simplest example would be the difference of resolution of NGS vs. microarray based transcriptomics). The virus-dependent, time-abstracted profiles then can be integrated into joint profiles using the same or similar formalisms as discussed in Sect. 4 and Lesne and Benecke (2008b)

probability density distributions are at the basis of the P-profiles generated from the to-be-annotated data. While so far only symmetric distributions have been described and studied (Lesne and Benecke 2008a, b), the formalism does not exclude the use of skewed, nontrivial distributions (Fig. 5). Furthermore, distance or divergence measures for skewed distributions, or parts thereof, can be defined. Thus instead of describing variability across individual measurements or different genetic backgrounds, the P_{pn} part of the probability annotation would capture a generalized evolution over time. In this manner, only a single profile would be created for the entire time-series where the actual number of measured discrete time points is replaced by a continuously modeled distribution. Those distributions then can be studied in a fashion similarly as to what has been briefly discussed in Sect. 4 and in more detail in Lesne and Benecke (2008b). Again, a number of different ways to achieve such integration have been proposed (Selinger 2012). Indeed, in the example of the respiratory virus infection (Sect. 3), the proposed integration mechanism provides a means of discerning which one is the more likely of the two possible mechanisms, and thus prioritize the experimentally testable hypotheses.

5 Concluding Remarks

Systems Biology is a rapidly evolving field with is receiving a great deal of attention in the field of infectious disease research owing to the potential to provide a greater understanding of the pathogen–host interactions that control infection phenotype and disease outcome. A key aspect of the systems approach is the use of computational methods to collectively integrate high-throughput omics and traditional virologic or histopathologic data into a systems-level view that allows the identification of functional processes involved in pathogen-associated disease and the further illumination of host targets representing key points of control by pathogens.

Albeit having already made strong arguments in favor of a systemic analysis of the pathogen, the host, and most importantly their joint, interdependent activity, taking these analyses to the next level will require to overcome many current conceptual, technical, statistical, and computational bottlenecks. A key aspect of a higher level understanding, linking objects and mechanisms to organs and phenotypes, will be the integration of data on the one hand, and inference of network structure and dynamics on the other, over multiple scales. This problem is far from trivial, and ideas of how it can be overcome are still rare and in the early stage of development.

The potentially defining role of the network dynamics of host–pathogen interactions, as discussed on two recent examples, exemplifies the urgent need of identifying solutions of how to handle time across scales. Based on a recent proposition of a probability-theory derived approach for functional genome representations a first glimpse of methodology that might turn out to handle at least some of the problems arising through time disparity over scales was developed. Obviously, this approach, and even more so generalizable ideas of overcoming scales, will need many iterations of scientific thought and experimentation before we will see major breakthroughs.

References

- Aderem A, Adkins JN, Ansong C, Galagan J, Kaiser S, Korth MJ, Law GL, McDermott JG, Proll SC, Rosenberger C, Schoolnik G, Katze MG (2011) A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm. *MBio* 2(1):e00325-10
- Bagnold RA (1936) The movement of desert sand. *Proc Royal Soc Lond A* 157(892):594–620
- Bécavin C, Barbi M, Victor JM, Lesne A (2010) Transcription within condensed chromatin: steric hindrance facilitates elongation. *Biophys J* 98:824–833
- Beigel JH, Farrar J, Han AM, Hayden FG, Hyer R, de Jong MD, Lochindarat S, Nguyen TK, Nguyen TH, Tran TH, Nicoll A, Touch S, Yuen KY (2005) Writing committee of the World Health Organization (WHO) consultation on human influenza A/H5. Avian influenza A (H5N1) infection in humans. *N Engl J Med* 353(13):1374–1385

- Belisle SE, Yin J, Shedlock DJ, Dai A, Yan J, Hirao L, Kutzler MA, Lewis MG, Andersen H, Lank SM, Karl JA, O'Connor DH, Khan A, Sardesai N, Chang J, Aicher L, Palermo RE, Weiner DB, Katze MG, Boyer J (2011) Long-term programming of antigen-specific immunity from gene expression signatures in the PBMC of rhesus macaques immunized with an SIV DNA vaccine. *PLoS One* 6(6):e19681
- Benecke A (2003) Genomic plasticity and information processing by transcriptional coregulators. *Com Plex Us* 1:65–76
- Benecke A (2006) Chromatin code, local non-equilibrium dynamics, and the emergence of transcription regulatory programs. *Eur Phys J E* 19:353–366
- Benecke A (2008) Gene regulatory network inference using out of equilibrium statistical mechanics. *HFSP J* 2:183–188
- Benecke A, Gale M Jr., Katze MG (2012) Dynamics of innate immunity are key to chronic immune activation in AIDS. *Curr Opin HIV & AIDS* 7(1):71–78
- Bosinger SE, Li Q, Gordon SN, Klatt NR, Duan L, Xu L, Francella N, Sidahmed A, Smith AJ, Cramer EM, Zeng M, Masopust D, Carlis JV, Ran L, Vanderford TH, Paiardini M, Isett RB, Baldwin DA, Else JG, Staprans SI, Silvestri G, Haase AT, Kelvin DJ (2009) Global genomic analysis reveals rapid control of a robust innate response in SIV-infected sooty mangabeys. *J Clin Invest* 119(12):3556–3572
- Bosinger SE, Sodora DL, Silvestri G (2011) Generalized immune activation and innate immune responses in simian immunodeficiency virus infection. *Curr Opin HIV AIDS* 6(5):411–418
- Bosinger SE, Jacquelin B, Benecke A, Silvestri G, Müller-Trutwin M (2012) Systems biology towards the understanding of nonpathogenic SIV infection in natural host primate species. *Curr Opin HIV AIDS* 7(1):71–78
- Bradel-Tretheway BG, Mattiacci JL, Krasnoselsky A, Stevenson C, Purdy D, Dewhurst S, Katze MG (2011) Comprehensive proteomic analysis of influenza virus polymerase complex reveals a novel association with mitochondrial proteins and RNA polymerase accessory factors. *J Virol* 85(17):8569–8581
- Brenchley JM, Silvestri G, Douek DC (2010) Nonprogressive and progressive primate immunodeficiency lentivirus infections. *Immunity* 32(6):737–742
- Bühlmann P, Wyner A (1999) Variable length markov chain. *Ann Stat* 27:480–513
- Cabal GG, Rodriguez-Navarro S, Genevesio A, Olivo-Marin JC, Zimmer C, Gadal O, Feuerbach-Fournier F, Lesne A, Buc H, Hurt EC, Nehrass U (2006) Molecular analysis of SAGA mediated nuclear pore gene gating activation in yeast. *Nature* 441:770–773
- Chang ST, Tchitchek N, Ghosh D, Benecke A, Katze MG A (2012) chemokine gene expression signature derived from meta-analysis predicts the pathogenicity of viral respiratory infections. *BMC Syst Biol* 5:202
- Cilloniz C, Pantin-Jackwood MJ, Ni C, Goodman AG, Peng X, Proll SC, Carter VS, Rosenzweig ER, Szretter KJ, Katz JM, Korth MJ, Swayne DE, Tumpey TM, Katze MG (2010) Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model of infection. *J Virol* 84(15):7613–7624
- Donnelly CA, Ghani AC, Leung GM, Hedley AJ, Fraser C, Riley S, Laith A, Abu-Raddad J, Ho LM, Thach TQ, Chau P, Chan KP, Lam TH, Tse LY, Tsang T, Liu SH, Kong JHB, Lau EMC, Ferguson NM, Anderson RM (2003) Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 361:1761–1766
- Ein-Dor L, Kela I, Getz G, Givol D, Domany E (2005) Outcome signature genes in breast cancer: is there a unique set?. *Bioinformatics* 21(2):171–178
- Estes JD, Gordon SN, Zeng M, Chahroudi AM, Dunham RM, Staprans SI, Reilly CS, Silvestri G, Haase AT (2008) Early resolution of acute immune activation and induction of PD-1 in SIV-infected sooty mangabeys distinguishes nonpathogenic from pathogenic infection in rhesus macaques. *J Immunol* 180(10):6798–6807
- Favre D, Lederer S, Kanwar B, Ma ZM, Proll S, Kasakow Z, Mold J, Swainson L, Barbour JD, Baskin CR, Palermo R, Pandrea I, Miller CJ, Katze MG, McCune JM (2009) Critical loss of the balance between Th17 and T regulatory cell populations in pathogenic SIV infection. *PLoS Pathog* 5(2):e1000295

- Goodman AG, Tanner BC, Chang ST, Esteban M, Katze MG (2011) Virus infection rapidly activates the P58(IPK) pathway, delaying peak kinase activation to enhance viral replication. *Virology* 417(1):27–36
- Gregorius HR (2006) The isolation principle of clustering: structural characteristics and implementation. *Acta Biotheor* 54(3):219–233
- Harris LD, Tabb B, Sodora DL, Paiardini M, Klatt NR, Douek DC, Silvestri G, Müller-Trutwin M, Vasile-Pandrea I, Apetrei C, Hirsch V, Lifson J, Brenchley JM, Estes JD (2010) Downregulation of robust acute type I interferon responses distinguishes nonpathogenic simian immunodeficiency virus (SIV) infection of natural hosts from pathogenic SIV infection of rhesus macaques. *J Virol* 84(15):7886–7891
- Hopfield JJ (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A* 71(10):4135–4139
- Hopfield JJ, Yamane T, Yue V, Coutts SM (1976) Direct experimental evidence for kinetic proofreading in amino acylation of tRNA^{Ala}. *Proc Natl Acad Sci U S A* 73(4):1164–1168
- Israeli N, Goldenfeld N (2006) Coarse-graining of cellular automata, emergence, and the predictability of complex systems. *Phys Rev E* 73:026203
- Jacquelin B, Mayau V, Targat B, Liovat AS, Kunkel D, Petitjean G, Dillies MA, Roques P, Butor C, Silvestri G, Giavedoni LD, Lebon P, Barr P, Sinoussi F, Benecke A, M'Yller-Trutwin MC (2009) Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. *J Clin Invest* 119(12):3544–3555
- Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6(6):451–464
- Kash JC, Basler CF, Garc'a-Sastre A, Carter V, Billharz R, Swayne DE, Przygodzki RM, Taubenberger JK, Katze MG, Tumpey TM (2004) Global host immune response: pathogenesis and transcriptional profiling of type A influenza viruses expressing the hemagglutinin and neuraminidase genes from the 1918 pandemic virus. *J Virol* 78(17):9499–9511
- Kullback S, Leibler R (1951) On information and sufficiency. *Ann Math Stat* 22:79–86
- Lavelle C, Benecke A (2006) Chromatin physics: replacing multiple, representation-centered descriptions at discrete scales by a continuous function-dependent selfscaled model. *Eur Phys J E (Soft Matter)* 19:379–384
- Lederer S, Favre D, Walters KA, Proll S, Kanwar B, Kasakow Z, Baskin CR, Palermo R, McCune JM, Katze MG (2009) Transcriptional profiling in pathogenic and non-pathogenic SIV infections reveals significant distinctions in kinetics and tissue compartmentalization. *PLoS Pathog* 5(2):e1000296
- Lepelley A, Louis S, Sourisseau M, Law HK, Pothlichet J, Schilte C, Chaperot L, Plumas J, Randall RE, Si-Tahar M, Mammano F, Albert ML, Schwartz O (2011) Innate sensing of HIV-infected cells. *PLoS Pathog* 7(2):e1001284
- Lesne A (1998) Renormalization methods. Wiley, Chichester, ISBN 0-471-96689-4
- Lesne A (2011) Morphogenesis. Springer, ISBN 978-3-642-13173-8
- Lesne A, Benecke A (2008a) Probability landscapes for integrative genomics. *Theor Biol Med Mod* 5:9
- Lesne A, Benecke A (2008b) Feature context-dependency and complexity reduction in probability landscapes for integrative genomics. *Theor Biol Med Mod* 5:21
- Lesne A, Lagües M (2012) Scale invariance. Springer, ISBN 978-3-642-15122-4
- Lesne A, Victor JM (2006) Chromatin fiber functional organization: some plausible models. *Eur Phys J E* 19:279–290
- Liu HM, Gale M (2010) Hepatitis C Virus Evasion from RIG-I-Dependent Hepatic Innate Immunity. *Gastroenterol Res Pract.* 2010:548390
- Loo YM, Gale M Jr. (2011) Immune signaling by RIG-I-like receptors. *Immunity* 34(5):680–692
- Manches O, Bhardwaj N (2009) Resolution of immune activation defines nonpathogenic SIV infection. *J Clin Invest* 119(12):3512–3515

- Maubourguet N, Lesne A, Changeux JP, Maskos U, Faure P (2008) Behavioral sequence analysis reveals a novel role for beta2* nicotinic receptors in exploration. *PLoS Comput Biol* 4(11):e1000229
- McKeithan TW (1995) Kinetic proofreading in T-cell receptor signal transduction. *Proc Natl Acad Sci U S A* 92(11):5042–5046
- Mir KD, Gasper MA, Sundaravaradan V, Sodora DL (2011) SIV infection in natural hosts: resolution of immune activation during the acute-to-chronic transition phase. *Microbes Infect* 13(1):14–24
- Moore C (1990) Unpredictability and undecidability in dynamical systems. *Phys Rev Lett* 64:2354–2357
- Pandrea I, Gaufin T, Gautam R, Kristoff J, Mandell D, Montefiori D, Keele BF, Ribeiro RM, Veazey RS, Apetrei C (2011) Functional cure of SIVagm infection in rhesus macaques results in complete recovery of CD4+ T cells and is reverted by CD8+ cell depletion. *PLoS Pathog* 7(8):e1002170
- Ross AL, Brave A, Scarlatti G, Manrique A, Buonaguro L (2010) Progress towards development of an HIV vaccine: report of the AIDS vaccine 2009 conference. *Lancet Infect Dis* 10(5):305–316
- Rotger M, Dalmau J, Rauch A, McLaren P, Bosinger SE, Martinez R, Sandler NG, Roque A, Liebner J, Bategay M, Bernasconi E, Descombes P, Erkizia I, Fellay J, Hirschel B, Mir JM, Palou E, Hoffmann M, Massanella M, Blanco J, Woods M, GYnthard HF, de Bakker P, Douek DC, Silvestri G, Martinez-Picado J, Telenti A (2011) Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. *J Clin Invest* 121(6):2391–2400
- Schreiber F, Lynn DJ, Houston A, Peters J, Mwafurirwa G, Finlay BB, Brinkman FS, Hancock RE, Heyderman RS, Dougan G, Gordon MA (2011) The human transcriptome during nontyphoid Salmonella and HIV coinfection reveals attenuated NF κ B-mediated inflammation and persistent cell cycle disruption. *J Infect Dis* 204(8):1237–1245
- Selinger C (2012) On diffusion processes arising from optimal transport with applications to negative selection. *Ann Math* (to appear)
- Smet-Nocca C, Paldi A, Benecke A (2010) From epigenomic to morphogenetic emergence. In: Bourguin P, Lesne A (eds) *Morphogenesis*. Springer, ISBN-13: 978-3-642-13173-8
- Sonnenschein N, Geertz M, Muskhelishvili G, Hÿtt MT (2011) Analog regulation of metabolic demand. *BMC Syst Biol* 5:40
- Spector DL (2003) The dynamics of chromosome organization and gene regulation. *Ann Rev Biochem* 72:573–608
- Suthar MS, Ma DY, Thomas S, Lund JM, Zhang N, Daffis S, Rudensky AY, Bevan MJ, Clark EA, Kaja MK, Diamond MS, Gale M Jr (2010) IPS-1 is essential for the control of West Nile virus infection and immunity. *PLoS Pathog* 6(2):e1000757
- Tisoncik JR, Katze MG (2010) What is systems biology?. *Future Microbiol* 5(2):139–141
- Tisoncik JR, Belisle SE, Diamond DL, Korth MJ, Katze MG (2009) Is systems biology the key to preventing the next pandemic?. *Future Virol* 4(6):553–561
- Widom J (1998) Structure, dynamics and function of chromatin in vitro. *Annu Rev Biophys Biomol Struct* 27:285–327
- Ye J, Maniatis T (2011) Negative regulation of interferon- β gene expression during acute and persistent virus infections. *PLoS One* 6(6):e20681