Neither Hide Nor Hair: The Difficulty of Identifying Useful Disease Biomarkers

See "Multigene Analysis Can Discriminate Between Ulcerative Colitis, Crohn's Disease, and Irritable Bowel Syndrome," by von Stein et al on page 1869.

The Identification of Agents of Disease. Robert Koch was the first to propose a series of definitive tests to determine the agents of infectious disease. Koch's postulates, first published in 1890, were subsequently updated for the molecular era in 1988 by Stanley Falkow[1] and continue to form the cornerstone of molecular pathogenesis. Indeed, detection of pathogen-associated molecules are widely used in diagnostic clinical tests,[2–4] especially in cases where alternative methods are both slow and expensive. However, the causative agents for many common, noninfectious diseases have proved far more difficult to identify than those of the infectious diseases, even when the full force of modern genomics and molecular medicine is brought to bear. This problem reflects inherent difficulties of complex disease diagnosis and classification— diseases may have common symptoms, but many diverse causes; diseases may have multiple, complex, interacting causes, not amenable to direct identification; or despite well-characterized symptoms, the underlying pathology may be poorly understood.

Thus far, despite the completion of the human genome sequence and resulting projects, such as HapMap and large-scale genome-wide association (GWA) studies for many complex diseases, molecular diagnostics is still in its infancy. Rationally designed treatments remain largely confined to small, well-studied areas of pathology, including leukemia (imatinib mesylate [Gleevec]), lung cancer (gefitinib [Iressa]), and autoimmune diseases (etanercept [Enbrel]). Indeed, a recent analysis of known drugs and their targets revealed that a relatively small number of cellular proteins are targeted by a disproportionate number of different drugs.[5] Partially, this is due to a "follow-on" effect where successful therapies are joined by similar drugs with improved properties and/or intellectual property profiles. However, it may also be the case that heavily drugged targets are highly amenable to therapy, either because of their accessibility, or strong association with the disease pathology, although because so little is known about the molecular characteristics of many drug targets and their relation to disease, the extent of this constraint is largely unknown.

The Difficulty of Diagnosis. Currently, the majority of disease diagnosis, and even treatment selection, is still accomplished via symptomatic, empirical physician assessment, combined with largely nonmolecular laboratory tests for pathology and phenotypic classification. In some settings, such methods have proved to be more accurate than those of molecular diagnosis—for example, differentiation of gastrointestinal tuberculosis and Crohn's disease (CD).[6] Although the established clinical methods have historically proved extremely successful and remain the gold standard against which molecular tests will be compared, in some areas they fall short of the ideal, particularly in identification of preclinical disease, unambiguous disease diagnosis, and rational treatment selection. Therefore, disease treatment often consists of a succession of more or less successful therapeutic regimens, until 1 is selected as a reasonable balance between remediation of disease symptoms and resultant side effects or other constraints. It is the hope of molecular medicine that this process could be circumvented, with disease diagnosis and treatment determined rapidly and unequivocally via specific and sensitive tests.

Nowhere has this apparent dichotomy of increased understanding of molecular risk factors and basis for disease, combined with little apparent improvement in diagnostic methods or clinical outcome, been more stark than for inflammatory bowel diseases (IBD). Genetic risk factors such as family history and specific variants in the *CARD15* gene have been identified for >7 years, and yet their role in pathology remains largely opaque.[7,8] Recent genome-wide screens for susceptibility have identified a large number of strongly associated gene variants for different IBD subtypes, as well as provided potential insight into the pathways underlying the disorders, such as autophagy and phagocytosis.[9–13] Some light has been shed onto treatment selection by the identification of the role of genetic variants of thiopurine *S*-methyltransferase on the processing of azathioprine and mercaptopurine, and genotyping is routinely used to guide dosage in these cases (see Teml et al[14] for a recent review). This remains a signal illustration of the role of pharmacogenetics in modern medicine, all the more striking because of the rarity of such examples.

Finding Diagnostic Disease Predictors. All molecular-based analyses require the identification of appropriate molecular targets that are both readily and sensitively assayable and specific to the disease or diseases under investigation. In the case of noninfectious diseases, these markers are likely to represent host responses to disease, rather than exogenous pathogen-derived molecules. The role of the host–pathogen interface in gener-

ating diverse disease phenotypes has been highlighted by recent studies in other diseases such as malaria,[15] and similar interactions are likely to play a role in IBD. However, in IBD the key factors for disease progression are likely to be the interplay between the host and the abundant bacteria in the gut, rather than a single pathogen. This represents an additional challenge to specificity, because many host and flora-derived molecules are likely to be involved in the pathology, thereby resulting in diverse disease states. Finding disease-specific biomarkers among the vast array of potential targets and confounding factors is therefore a large part of the challenge to development of molecular diagnostics. This challenge is exacerbated by the fact that the most useful tests would be able to distinguish diseases with a high commonality of symptoms, such as gastrointestinal inflammation in the case of IBD.

In this issue of GASTROENTEROLOGY, von Stein et al[16] present a novel approach to identification of such diagnostic markers in the context of IBD. In an attempt to distinguish between CD, ulcerative colitis (UC), and irritable bowel syndrome (IBS), they isolated RNA from selected biopsy specimens from inflamed and noninflamed areas of colon from 8 UC patients. These samples were subjected to suppressive, subtractive hybridization (SSH), yielding a set of clones highly enriched for sequences overrepresented in inflamed tissues. The SSH technique relies upon the use of 2 cDNA populations, the driver (in this case derived from uninflamed tissues) and the tester (inflamed). The tester population is split into 2 pools (1 and 2) and each fraction is differentially tagged with a 5′ adapter. Each pool is each mixed with an excess of driver cDNAs and allowed to hybridize. In this step, the majority of transcripts present in both tester and driver samples form heterohybrids, containing both a tagged and an untagged strand. Those cDNAs abundant in the tester population tend to form homohybrids, whereas less abundant, tester-only cDNAs are enriched for single-stranded forms. This stage acts as a normalization step, because abundant cDNAs tend to homohybridize, with less abundant cDNAs favored for maintenance as single strands. The 2 differentially tagged cDNA pools are then mixed, with the addition of more denatured driver cDNA. Annealing then takes place only between those remaining single-stranded cDNAs that are not present in the driver pool, resulting in a subset of cDNA hybrids that have tags from both pools 1 and 2. These heterotagged hybrids are then amplified by tag-specific polymerase chain reaction (PCR), resulting in a vast enrichment of those cDNAs specific to the tester sample. In the case of von Stein et al,[16] 2 such subtractions were performed to allow identification of both up- and downregulated genes. Selection from these enriched pools and confirmation by quantitative PCR resulted in 7 candidate cDNAs highly differentially expressed in the inflamed tissues of UC and CD patients. Because these genes were differentially expressed in UC and CD patients, their usefulness as a diagnostic test was determined using quantitative RT-PCR profiling of endoscopically retrieved tissue samples. A pilot study showed that with these 7 biomarkers, it was possible to correctly classify UC or CD patients in >92% of cases, close to the abilities of conventional clinical diagnosis.

After this success, the same 7 biomarkers were used to classify 20 patients with unclassified IBD. Ten of these patients were subsequently able to be classified by an independent physician using standard clinical criteria. Of these 10 patients, 9 were correctly classified by the biomarker analysis. In the case of 8 patients subsequently diagnosed as having IBS, their noninflamed biopsy samples were all correctly classified as non-UC. Further refinement and testing of the diagnostic algorithm enabled highly accurate classification of IBD and non-IBD cases to UC, CD, or non-IBD groups. The success of this method seems to rely on the use of 7 biomarkers, each of which is reasonably discriminatory, but together yield a high level of specificity and sensitivity. It will be of great importance to fully validate these tests and algorithms in both the inpatient and outpatient settings. It will also be of interest if they can be adapted to monitor responses to therapy and indications of future relapse/remission.

Having undertaken such a broad, unbiased effort to find differentially expressed genes in UC and CD, von Stein et al[16] have also identified a set of proteins likely to be of interest in the broader disease context. Indeed, the solute carrier transport molecules identified have been previously observed to be dysregulated in IBD, possibly owing to the increased permeability and elevated epithelial turnover of the inflamed gut. SPAP is a PDZK1 interactor with a proposed role in proliferation and carcinogenesis, as well as reactive oxygen species production. PDZK1 itself is thought to play a role in the targeting of nitric oxide synthase 2 to the apical surface of epithelial cells and phagosomes, a process known to be disrupted by bacterial pathogens.[17,18] Two of the identified genes have a near-direct interaction—GRO-α is thought to bind to the cell-surface proteoglycan syndecan-1, matrix metalloproteinase 7 is able to cleave syndecan-1 from the cell surface, and the resulting chemokine/glycan complexes act as potent neutrophil chemoattractants.[19] Reg-IV has previously been identified as a potential marker of intestinal epithelial repair and regeneration, as well as carcinogenesis,[20,21] and is highly expressed in neuroendocrine cells, akin to the CD-associated gene *Phox2B*.[10] Finally, vanin-1 is thought to play a role in both protection against oxidative stress and induction of inflammation, with some studies suggesting that vanin-1 acts as a sensor of epithelial damage and thus exerts a dominant control over innate immune reactions in the gut.[22]

Extending the Phenotype. Although the application of SSH to find differential IBD biomarkers is novel,

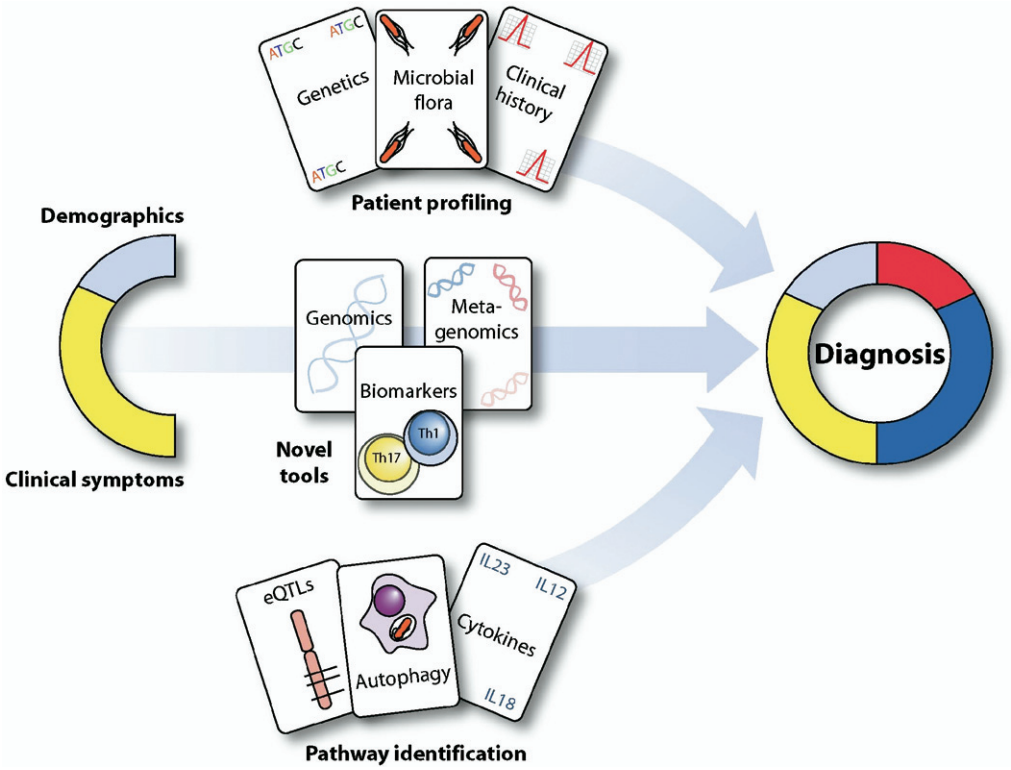

Figure 1. An idealized diagnostic scheme for the postgenomic era. Current diagnoses of IBD rely mostly on clinical symptoms, with a contribution from demographic data (such as smoking status, ancestry, and age). We have divided proposed molecular diagnostic avenues into 3 classes: patient profiling, novel tools, and pathway identification. Patient profiling would utilize a small number of molecular tests (such as multiplex PCR) for high-risk/highly differentiating genetic loci and aberrant microbial flora, as well as taking into account history of gastroenteritis and antibiotic-associated diarrhea. Novel tools encompasses the next generation of highly parallel, multiplexed tests for identified genetic and microbial risk factors or diagnostic biomarker patterns. Finally, identification of cellular and signaling pathways perturbed in IBD patients will generate novel probes to identify disease, possibly before clinical manifestation and yield novel avenues to treatment.

this is not the first time that biomarkers have been identified with the intent to distinguish between UC and CD. Using a combination of clinical assessment (including endoscopy) and inflammatory protein markers from both serum and feces, Langhorst et al[23] were able to accurately diagnose 95% of UC patients. However, this represents only a modest improvement over current methodologies and still relies on invasive endoscopic examination. An ideal test would allow a highly specific diagnosis using a minimally invasive sample, such as blood or stool. Recently, the speed and low cost of large-scale DNA sequencing has led to the development of "metagenomics," profiling of microbial communities by high-throughput shotgun sequencing (see Frank and Pace[24] for a recent review). It is possible that metagenomic analyses may result in useful differential diagnostics; the profiles of microbial communities seem to be profoundly altered in patients with inflammatory gastrointestinal disease.[25] Indeed, it has been shown that the composition of the gut flora can have profound effects on the host, and such flora seems to be both labile and responsive to factors such as diet and nutritional status.[26] Blood test panels of biomarkers do exist and are in clinical use; the IBDX panel from Glycominds is 1 such test, using 4 markers.[27] However, the discriminatory power of these tests is not yet comparable with current gold standard clinical evaluations and in unclassified IDB their discriminatory power is lower than the tests described by von Stein et al.[16] However, there have been several previous biomarker efforts whose early promise failed to be fulfilled—replication of these results in other cohorts will be essential before the true utility of the multigene analysis described by von Stein can be determined.

The real challenge in molecular diagnosis of IBD lies in leveraging the latest knowledge from GWA studies and molecular pathology to generate a large panel of biomarkers. Although the genes identified in such susceptibility studies are unlikely to be of use by themselves, identification of the pathways that they are involved in may provide insight. For example, there have been several reports of susceptibility to CD associated with interleukin (IL)-23R mutations. It is known that the IL-23 cascade contributes to the differentiation of the T helper (Th)17 subset of T cells, which play a key regulatory role in the gut. Recently, a population of Th17 cells was identified in CD patients, possessing both Th17 and Th1 features, suggesting that dysregulated Th17 differentia-

tion may occur in some CD patients.[28,29] One can envision that examination of surface markers (such as IL-23R) on circulating T cells, if performed with sufficient fidelity, might serve as a useful additional diagnostic tool.

The use of large panels of biomarkers is less likely to yield spurious results owing to patient–patient variation, provided the markers are well-chosen. The unbiased screening approach taken by von Stein et al[16] reflects the methods used in high-throughput genotyping and screening efforts and extension of these techniques is likely to yield further fruit. However, although extending the scope of biomarker panels will be useful, there may be difficulties in maintaining specificity and sensitivity when using such a large set of variables. Each marker will itself have to have very high specificity to ensure that noise does not increase, swamping gains achieved through increased biomarker depth. In addition, these biomarkers will have to be rigorously tested in healthy controls to ensure that specificity of diagnosis can be maintained in asymptomatic individuals. These large panels will also have to be robust enough to withstand a multitude of confounding factors; simultaneous infections, concurrent use of therapies and potential alternative causes of initial symptoms, such as cancer. All of these constraints will result in compromises, both to the effectiveness of molecular diagnosis and the invasiveness of the procedures required. For example, it could be argued that, compared with blood samples, the use of gut biopsies would lead to more robust IBD diagnoses, owing to the restriction of sampling to the anatomic site concerned, thus eliminating many confounding variables such as concurrent respiratory infection. However, this naïve view may prove incorrect; the very diversity of the gut flora and the maelstrom of inflammation may mask specific disease signals, such as abnormal T-cell subsets, which would be detectable in the peripheral circulation.

The compromises and trade-offs involved in generating sensitive, yet specific, tests for disease largely preclude the development of a single ideal test. However, our increasing knowledge of disease process and pathology is likely to yield increasing gains in early and robust diagnoses. In Figure 1, we present an idealized scheme for utilizing current and novel technologies, including patient profiles, molecular tools, and cellular pathway-based diagnostics as part of an overall drive toward rational diagnosis. We see the extension of diagnostics into the molecular realm adding to, not replacing, current gold standard clinical metrics and with increased understanding of the pathways affected, improving treatment selection. Large, molecular-based screening efforts such as undertaken by von Stein et al[16] are both promising to deliver more useful molecular diagnostic tools, but are also likely to further our knowledge of disease mechanisms and therefore greatly improve both diagnosis and therapy in the future.

ALAN HUETT
RAMNIK J. XAVIER
Gastrointestinal Unit and The Center for Computational and Integrative Biology
Massachusetts General Hospital and Harvard Medical School
Boston, Massachusetts

References

1. Falkow S. Molecular Koch's postulates applied to microbial pathogenicity. Rev Infect Dis 1988;10(Suppl 2):S274–S276.
2. Taegtmeyer M, Beeching NJ, Scott J, et al. The clinical impact of nucleic acid amplification tests on the diagnosis and management of tuberculosis in a British hospital. Thorax 2008;63:317–321.
3. Ashworth M, Horan KL, Freeman R, et al. Use of PCR-based Mycobacterium tuberculosis genotyping to prioritize tuberculosis outbreak control activities. J Clin Microbiol 2008;46:856–862.
4. Hoek PK, Gey van Pittius CN, Moolman-Smook H, et al. Fluorometric assay for susceptibility testing of rifampicin in mycobacterium tuberculosis complex (FAST-Rif). J Clin Microbiol 2008;46:1369–1373.
5. Yildirim MA, Goh KI, Cusick ME, et al. Drug-target network. Nat Biotechnol 2007;25:1119–1126.
6. Amarapurkar DN, Patel ND, Rane PS. Diagnosis of Crohn's disease in India where tuberculosis is widely prevalent. World J Gastroenterol 2008;14:741–746.
7. Kobayashi KS, Chamaillard M, Ogura Y, et al. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science 2005;307:731–734.
8. Maeda S, Hsu LC, Liu H, et al. Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. Science 2005;307:734–738.
9. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 2007;39:207–211.
10. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 2007;39:596–604.
11. Parkes M, Barrett JC, Prescott NJ, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 2007;39:830–832.
12. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.
13. Cho JH, Weaver CT. The genetics of inflammatory bowel disease. Gastroenterology 2007;133:1327–1339.
14. Teml A, Schaeffeler E, Herrlinger KR, et al. Thiopurine treatment in inflammatory bowel disease: clinical pharmacology and implication of pharmacogenetically guided dosing. Clin Pharmacokinet 2007;46:187–208.
15. Daily JP, Scanfeld D, Pochet N, et al. Distinct physiological states of Plasmodium falciparum in malaria-infected patients. Nature 2007;450:1091–1095.
16. von Stein P, Lofberg R, Kuznetsov NV, et al. Multigene analysis can discriminate between ulcerative colitis, Crohn's disease, and irritable bowel syndrome. Gastroenterology 2008;134:1869–1881.
17. Maresca M, Miller D, Quitard S, et al. Enteropathogenic *Escherichia coli* (EPEC) effector-mediated suppression of antimicrobial nitric oxide production in a small intestinal epithelial model system. Cell Microbiol 2005;7:1749–1762.
18. Davis AS, Vergne I, Master SS, et al. Mechanism of inducible nitric oxide synthase exclusion from mycobacterial phagosomes. PLoS Pathog 2007;3:e186.

19. Li Q, Park PW, Wilson CL, et al. Matrilysin shedding of syndecan-1 regulates chemokine mobilization and transepithelial efflux of neutrophils in acute lung injury. Cell 2002;111:635–646.
20. Kamarainen M, Heiskala K, Knuutila S, et al. RELP, a novel human REG-like protein with up-regulated expression in inflammatory and metaplastic gastrointestinal mucosa. Am J Pathol 2003;163:11–20.
21. Lu B, Xu J, Zhu Y, et al. Systemic analysis of the differential gene expression profile in a colonic adenoma-normal SSH library. Clin Chim Acta 2007;378:42–47.
22. Berruyer C, Pouyet L, Millet V, et al. Vanin-1 licenses inflammatory mediator production by gut epithelial cells and controls colitis by antagonizing peroxisome proliferator-activated receptor gamma activity. J Exp Med 2006;203:2817–2827.
23. Langhorst J, Elsenbruch S, Koelzer J, et al. Noninvasive markers in the assessment of intestinal inflammation in inflammatory bowel diseases: performance of fecal lactoferrin, calprotectin, and PMN-elastase, CRP, and clinical indices. Am J Gastroenterol 2008;103:162–9.
24. Frank DN, Pace NR. Gastrointestinal microbiology enters the metagenomics era. Curr Opin Gastroenterol 2008;24:4–10.
25. Frank DN, St Amand AL, Feldman RA, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 2007;104:13780–13785.
26. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 2006;444:1027–1031.
27. Dotan I, Fishman S, Dgani Y, et al. Antibodies against laminaribioside and chitobioside are novel serologic markers in Crohn's disease. Gastroenterology 2006;131:366–378.
28. Annunziato F, Cosmi L, Santarlasci V, et al. Phenotypic and functional features of human Th17 cells. J Exp Med 2007;204:1849–1861.
29. van Beelen AJ, Zelinkova Z, Taanman-Kueter EW, et al. Stimulation of the intracellular bacterial sensor NOD2 programs dendritic cells to promote interleukin-17 production in human memory T cells. Immunity 2007;27:660–669.

Address requests for reprints to: Alan Huett, The Center for Computational and Integrative Biology, Massachusetts General Hospital, Richard B. Simches Research Center, 185 Cambridge Street, Boston, MA 02114. e-mail: ahuett@ccib.mgh.harvard.edu

0016-5085/08/$34.00
doi:10.1053/j.gastro.2008.04.013

Unraveling the Role of PD-1/PD-L Interactions in Persistent Hepatotropic Infections: Potential for Therapeutic Application?

See "Functional Restoration of HCV-Specific CD8 T Cells by PD-1 Blockade is Defined by PD-1 Expression and Compartmentalization," by Nakamoto N, Kaplan DE, Coleclough J, et al, on page 1927; and "Dynamic Programmed Death 1 Expression by Virus-Specific CD8 T Cells Correlates With the Outcome of Acute Hepatitis B," by Zhang Z, Zhang JY, Wherry EJ, et al, on page 1938.

Chronic viral infections are an enormous problem worldwide, with >500 million people infected with human immunodeficiency virus (HIV), hepatitis B virus (HBV), or hepatitis C virus (HCV). There is no cure for HIV infection or for most patients with HBV and HCV infections, and preventive vaccines for HIV and HCV may be decades away. Studies of the adaptive immune responses to these viruses have demonstrated that both the $CD4^+$ and $CD8^+$ T-cell responses are important for viral control and for mediating clearance of HBV and HCV in the minority of patients that are able to spontaneously clear their infection (reviewed in Bowen and Walker[1]). However, in the majority of patients in whom virus persists, antigen-specific T cells have impaired effector function. These cells have been described as being "exhausted" owing to persistent antigen stimulation and have impaired ability to proliferate and to produce cytokines such as interleukin-2, tumor necrosis factor-α and interferon (IFN)-γ.

Recently, a molecule in the CD28 family of receptors, programmed death 1 (PD-1), was shown to be a marker of these functionally exhausted T cells.[2] Binding of PD-1 to one of its ligands, PD-L1 or PD-L2, transmits a negative signal to the T cells expressing PD-1, reducing cytokine production and proliferation (reviewed in Keir et al[3]). The PD-1/PD-L system may serve to modulate immune responses and to quell potentially harmful or overzealous T cells, and yet, pathogens may have evolved to seize this pathway for their own benefit. In mice chronically infected with lymphochoriomeningitis virus, blockade of the PD-1/PD-L1 interaction resulted in improved viral control and clearance of virus from several organs.[2] This improved viral control correlated with improved functionality of antigen-specific $CD8^+$ T cells.[2] As such, these initial studies demonstrated that manipulation of the PD-1/PD-L1 co-inhibitory system may have important therapeutic potential for chronic viral infections. This inhibitory system has particular relevance for hepatotropic viral infections because PD-L1 is highly expressed in the liver on mouse sinusoidal endothelial cells and Kupffer cells, and these cells have been shown to be capable of inhibiting proliferation of PD-1 expressing effec-