



Published in final edited form as:

*Clin Pharmacol Ther.* 2011 August ; 90(2): 323–327. doi:10.1038/clpt.2011.120.

## Applications of Translational Bioinformatics in Transplantation

Purvesh Khatri<sup>1,2</sup>, Minnie M Sarwal<sup>2,3</sup>, and Atul J Butte<sup>1,2</sup>

<sup>1</sup>Division of Systems Medicine, Department of Pediatrics, Stanford University, Stanford, CA

<sup>2</sup>Lucile Packard Children's Hospital, Stanford, CA

<sup>3</sup>Department of Pediatrics, Stanford University, Stanford, California, United States of America

### Introduction

Biomedical research is increasingly enamored with the promise of translational medicine. In an era of increasing costs and difficulties in drug discovery, the importance is now heightened for bioinformatics to advance and make more efficient this process of translational medicine (1). Here we discuss the potential role for translational bioinformatics to address current and future needs specifically in the field of transplantation, a field with continued unmet diagnostic and therapeutic needs.

### Emergence of Translational Bioinformatics

The editorial board of the recently launched *Journal of Translational Medicine (JTM)* recently confessed that they are baffled by reviewers' comments when they dismiss translational manuscripts with comments such as "not hypothesis-driven" or "not mechanistic" (2). The confession from the editorial board of *JTM* underscores the general acknowledgement that traditional "hypothesis-driven research alone cannot meet the needs of translational medicine"(2).

This is clearly evident as the number of Food and Drug Administration (FDA) approved drugs has been relatively constant to about 20 drugs per year, except for a brief increase in the mid 1990's as a consequence of the Prescription Drug User Fee Act (PDUFA) in 1992 (3), amid an increase in the cost of drug discovery from \$138 million in 1975 to \$1.3 billion in 2006 and more than 15 years needed on average in developing a single drug (4). In fact, an analysis of more than 1200 drugs and molecules approved by FDA since 1950 showed that the rate of new drug production of a pharmaceutical company follows a Poisson distribution and is constant (about 2–3 drugs per year at most). This constant rate of output is often blamed on the traditional hypothesis-driven research model, primarily because hypotheses derived from complex experimental models often do not translate to human pathology. On the other hand, we must also keep in mind that the research model is one of the factors affecting the drug approval rates. For instance, drug approval rates are also affected by the regulatory process, which is independent of a research model.

Nowhere is this problem more acute than in the field of transplantation (Fig. 1). Although short-term survival rates of grafts have increased, long-term graft survival rates have not changed much (5). Five year graft survival for transplanted organs varies from 43% for lung to 78% for kidney, which highlights the need for better understanding of post-transplant

injury mechanisms. In addition, there are few available non-invasive diagnostic tests for monitoring the long-term care of transplant patients.

In the last two decades, a number of high throughput technologies that enable simultaneous quantification of molecular states of tens of thousands of genes and proteins inexpensively have been made available. These technologies are constantly improving the resolution and coverage of molecular profiling by allowing identification of every Single Nucleotide Polymorphism (SNP) and transcript in a genome. Furthermore, the cost of these high throughput technologies is rapidly falling. For instance, using the next-generation sequencing a whole genome is expected to be less than \$1000 in the near future from over \$30,000 per genome.

In parallel with the explosion of molecular measurements generated by high throughput technologies, a fundamental change has occurred in the way this molecular data is shared. Many journals, especially high-impact journals, require the data to be submitted to international public databases before a manuscript is considered for publication. A number of international public data repositories such as the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and the European Bioinformatics Institute (EBI) Array Express have been set up to store and distribute these data. Interestingly, this wave of public data sharing has not been rapidly embraced by the transplant field as evidenced by many of the transplant related journals still not mandating raw data deposition in the public domain, prior to acceptance for publication.

The importance of sharing large molecular data sets across experiments cannot be overestimated; as of January 2011, GEO contains more than a half million samples from over 21,000 experiments. Furthermore, these databases are complemented by knowledge-bases of functional annotation that describe the biological processes and signaling pathways in which these gene products are known to be involved in. Integration of this knowledge and data presents unprecedented opportunities that hold a promise to accelerate and improve our understanding of biology, and specifically for the field of transplantation.

## High throughput molecular measurements in transplantation

By far, gene expression microarrays have been the most used high throughput technology in transplantation to date (6). Despite delayed adoption of the microarray technology in transplantation, the number of studies using high throughput technologies has been steadily increasing, albeit slowly (7). Since the publication of the first large transplant microarray study in 2003 (microarrays were invented in 1995), there are now over 70 human studies in public domain using high throughput technologies. These studies examined biopsy, blood and urine samples from different conditions including acute rejection (AR), stable graft functions (STA), chronic injury (CI), tolerance (TOL), and drug response (DR) from transplant patients.

Increased use of high throughput technologies has led to significant improvements in our understanding of the complex allograft injury mechanisms. In a landmark study, Sarwal et al., identified a pivotal role of infiltrating B-cells in acute rejection demonstrating a strong association between presence of dense clusters of B-cells and severe acute rejection (8). Similar pathogenesis-based transcripts (PBT) expression panels have been inferred from mouse experiments and applied to human transplant expression patterns in an effort to develop correlates of histopathological lesions in renal transplant biopsies (9). Most recently, Pham, Valentine, and colleagues from the IMAGE Study Group showed in a landmark study how blood-based gene-expression profiling could be used to substitute for biopsy based monitoring after heart transplantation (10).

Newer modalities have enabled studies to move past RNA into their coded proteins. To evaluate pathogenicity of non-HLA antibodies after transplantation, Sutherland et al. used newer protein microarrays to identify 36 non-HLA targets in multiple renal transplant patients with acute renal transplant rejection. From this list, protein Kinase C- $\zeta$  (PKC $\zeta$ ) was then validated to show that it is a marker of severe allograft injury (11). Additionally, using protein microarrays Angiotensinogen and PRKRIP1 have been identified as biomarkers of chronic kidney injury, and it is hypothesized that autoantibodies are raised against these unusual targets as they are exposed in the process of cellular damage in the kidney (12).

Similar to the progress seen in genomics and proteomics, recent advances in small molecule identification technologies (e.g., mass spectrometry, surface enhanced laser desorption/ionization, Liquid Chromatography/Mass Spectrometry, nuclear magnetic resonance) have given rise to the application of peptidomics and metabolomics to transplantation. Urine is a rich biofluid source for biomarker discovery in organ transplantation. Shotgun proteomics can now map the entire urinary proteome (13), and evaluate its perturbation during different types of graft injury. Smaller fragments of the urinary peptidome, consisting of degraded byproducts of intact proteins by enzymatic cleavage, can also provide insights into the perturbations in chemical balance during kidney injury (14–15). Metabolomics has been used for identifying injury caused by ischemia and reperfusion injury (16) as well as for monitoring drug toxicity (17–18). Metabolomics may be more ideally suited for monitoring drug toxicity than other high throughput technologies, as small molecule drugs and drug metabolites can be specifically measured (19).

High throughput technologies have also been expanded to study the role of recently discovered biological entities in transplantation, such as microRNAs (miRNAs). Anglicheau et al. recently used microfluidic cards to profile miRNAs in post-transplantation biopsies to demonstrate their altered expression during AR (20). Furthermore, they also demonstrated that the miRNAs over-expressed in AR biopsies are also highly expressed in peripheral blood mononuclear cells (PBMCs), which suggested that the intragraft change in miRNA levels may be explained by infiltrating cells, hence, may be used as potential non-invasive biomarkers.

## **Current challenges in transplantation and the role of translational bioinformatics in overcoming them**

However, despite these advances in our understanding of graft injury mechanisms, the impact on diagnostics and therapeutics for transplantation has been limited, as evident from the excellent short-term graft survival being overshadowed by poor long-term survival, the need for life-long immunosuppressive medication, and lack of non-invasive markers for monitoring and predicting graft injury.

### **Small sample sizes**

One of the reasons for the limited impact of the high throughput studies in transplantation has been typically lower number of individuals and samples used these studies. For instance, searching the NCBI GEO for microarray studies in humans described with the term “transplant” yields 69 experiments, of which only 16 have more than 50 samples and only 6 have more than 100 samples. These numbers are more disappointing when put into the context that these experiments are divided among four different organs (lung, kidney, liver, heart) studying at least three different conditions (acute rejection, chronic rejection, tolerance). In other words, since the adoption of high throughput technologies in transplantation almost a decade ago, there have still not been enough studies with a large enough number of samples to truly understand graft injury mechanisms to bring novel

diagnostics and therapeutics into the clinics that can improve patient care. One way to address this shortcoming is to perform a meta-analysis by integrating smaller independent experiments. Such an analysis can not only increase the number of available samples, but also account for the experiment-specific technical biases such as microarray platform or hybridization protocol (21)(22).

In one example of an integrative analysis, Chen et al. recently performed a meta-analysis using three transplant RNA microarray data sets from AR biopsies (two from kidney and one from heart transplant) (23). Using a vote counting method, 45 genes were identified that were significantly over-expressed in all data sets. The proteins coded for by these RNA were then screened as potential blood markers for AR, of which three proteins (PECAM1, CD44, and CXCL9) were found to be significantly over-expressed in blood samples in both kidney and heart transplant patients with biopsy-confirmed AR. One of the markers, PECAM1, identified renal AR with 89% sensitivity and 75% specificity, and had an area under the receiver operating characteristic (ROC) curve of 0.716 for cardiac transplant patients. This study demonstrated that integration of data sets can reduce biological biases across experiments, even the effect of tissue source. In addition to reducing technical and biological biases, integration of public data sets from different organ transplants allows positing and testing novel hypotheses such as identifying common immune responses irrespective of the type of transplanted organ (24).

Integration can also be performed across molecular measurements types. Li et al. integrated antibody-level measurements from a protein array with renal compartment-specific gene expression data (25). This integrative analysis showed that some of the post-transplant serological responses observed using protein microarrays were specific to the transplanted organ.

### Diagnostic and therapeutic needs in transplantation

Another challenge in transplantation is the lifelong administration of current immunosuppressive drugs with multiple side effects. For instance, use of calcineurin inhibitors itself is associated with nephrotoxicity, which in turn can contribute to long-term graft failure, along with opportunistic infections. Similarly use of corticosteroids increases the risk of cardiovascular diseases. Translational bioinformatics could play a significant role in addressing the critical need of identifying new immunosuppressive targets in transplantation.

Virtually all existing immunosuppressives are designed to prevent acute rejection by inhibiting T-cell activation. This is achieved through various pathways, such as inhibition of antigen presenting cell development, cytokine production, or co-stimulatory signals for T cell activation. However, reduction in AR incidence has been shown to have minimal or no effect on graft survival (26). Furthermore, existing drugs are unable to prevent chronic rejection. Chronic rejection is thought to be caused by alloantigen-independent mechanism in addition to alloantigen-dependent mechanisms.

The development of additional therapeutic options for transplantation may benefit from a systems-level approach. As an example, when TGN1412, an anti-CD28 monoclonal antibody, was administered to six patients in a phase 1 clinical trial in 2006, all patients suffered from severe multi-organ failure within hours following the induction of pro-inflammatory cytokines (27). CD28 is a co-stimulatory receptor on T cell, which binds to CD80 or CD86 on activated antigen-presenting cells. Inhibiting interaction between CD28 and CD80/CD86 has been shown to inhibit a variety of immune responses in vivo, including transplant rejection. Although the authors did not report the mechanism for the release of cytokine storm, Puellmann and Beham hypothesized that TGN1412 may also have activated

neutrophils as a subset of human neutrophils also expresses both CD28 and T cell receptor (28). These undesired side effects of TGN1412 highlight the limitations of focusing on single pathway in predicting a system-level response to an external stimulus, as they may have been potentially avoided by a global analysis.

Allograft rejection is a heterogeneous process starting with antigen processing and presentation, followed by cytokine-cytokine receptor signaling and activation of different immune cells that ultimately lead to graft failure. Hence, designing a drug that targets single pathway (e.g., co-stimulatory blockade of T cell activation) in isolation is, intuitively, unlikely to improve allograft survival in the long term. In other words, the TGN1412 trial under scores the need for studying complex immunological systems, such as allograft injury, at a systems biology level, especially since there are multiple known mechanisms of graft injury.

### **Scientists enabled by systems biology will drive future transplantation research**

Systems biology expands from the traditional molecular biological method of studying pairwise interactions into a network-based approach by integrating individual components to model a complex system. These individual molecular relationships can be built from a variety of components (29). By integrating data of various types, systems biology aims to explain a disease at the level of regulatory pathways in tissues and organs, even in whole organisms, while attempting to account for dynamics within regulatory networks. A large number of computational approaches have been developed to generate co-expression networks from protein binding data (30), functional annotations (31) and drug activity (32). Using these approaches it has been shown that such networks have properties that are not otherwise discernable from the relations themselves, and have preferential connectivity that results in “hub” nodes, which are molecules that connect to a larger number of other molecules (33). These hubs have been shown to be critical in a number of studies (34–35).

There have been strong arguments for using systems biology-based techniques for identifying critical component nodes in order to improve drug discovery (36–37). However, to the best of our knowledge, the use of systems biology-based approaches in transplantation has been very limited. To date, a large number of biomarkers have been identified for various post-transplant conditions without sufficient evidence discussing whether these are markers of an ongoing injury (effect markers) or related to the actual causes of the injury (causal markers). Although tremendously useful in graft monitoring, effect markers cannot be directly used to prevent injury, and are not useful as targets for drug development. Development of new drugs that reduce drug toxicity and chronic rejection requires identification of causal markers that can be targeted for novel therapeutics. We believe that use of systems biology is the critical next step for deeper understanding and the identification of causal markers of graft injury in transplantation.

However, integration of large amounts of data for systems biology-based analysis poses new challenges for the field of transplantation. Note as an example that a typical microarray experiment produces approximately 50,000 data points per sample. Hence, an experiment with 50 samples will produce more than 2.5 million data points. This number is dwarfed when compared to millions of data points generated by SNP genotyping platforms for thousands of samples in a typical genome-wide association study. Furthermore, the amount of data generated is only bound to increase as next-generation sequencing becomes more commonly used.

Therefore, a future scientist in transplantation using systems biology-based approaches will now be required to have training in multiple disciplines, including computational biology as well as the traditional clinical sciences. For instance, a clinical scientist evaluating data from

a high throughput technology, using any of the methods available for analysis of the data, needs to be aware of the nuanced differences between the various methods and their effects on interpretations of the data.

What this means is that today's clinical scientist is going to need a basic, if not advanced, computer programming ability. A clinician scientist must be able to integrate these data quickly and in meaningful way. A relevant example was recently noted in the leaked emails of a Climatic Research Unit (CRU) employee at the University of East Anglia in the UK. The employee wrote in his notes "Yup, my awful programming strikes again," as he tried to correct code for analyzing weather station data (38). Without necessary programming skills, it is easy to imagine unexpected consequences that can cast doubts on the results from the entire field. At the same time, incorporation of computational skills into the curricula of transplantation training program is not yet a high priority, to our knowledge.

## Conclusion

Since the first use of microarrays in transplantation almost a decade ago, the use of high throughput technologies in transplantation has significantly improved and advanced our understanding of allograft rejection mechanisms. However, although short-term survival rates have been excellent, long-term survival rates for transplanted organs have not improved, and drug toxicity and chronic rejection remain major challenges.

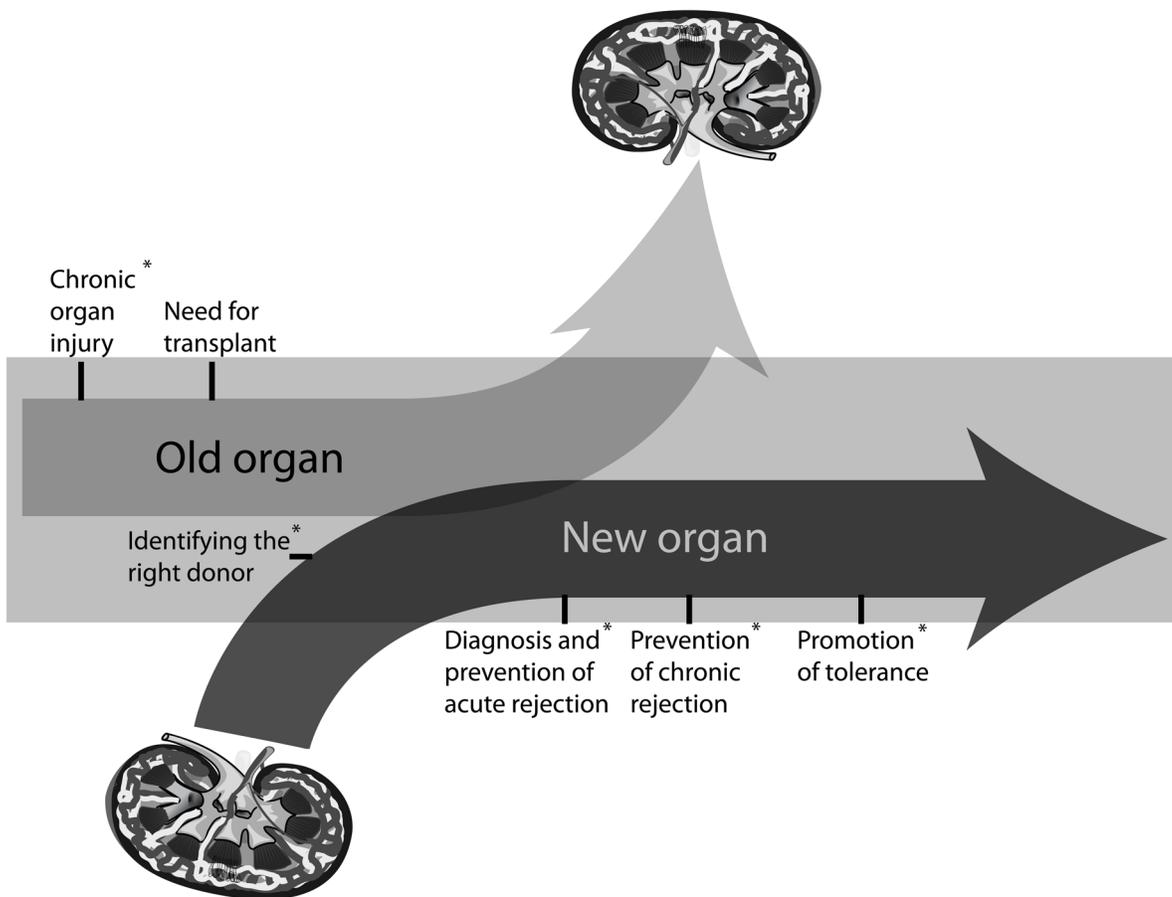
In order to take our understanding of injury mechanisms in organ transplantations to the next level, integration of molecular measurement data from different experiments and different technologies is required. Furthermore, these integrated data need to be analyzed at a global, systems biology level, to identify better diagnostic and therapeutic markers. However, the fragmented and incomplete nature of the existing knowledge bases still poses a challenge to achieving these goals (39). Furthermore, wider adoption of a policy to submit raw data into public repository is also required by the transplant related journals. It is also imperative that the next generation of clinician scientists is armed with computational skills that will ensure novel questions continued to be posed and answered, enabled by the proper integration of diverse sources of data.

## References

1. Butte AJ. Translational bioinformatics: coming of age. *J Am Med Inform Assoc.* 2008; 15:709–14. [PubMed: 18755990]
2. Mankoff SP, Brander C, Ferrone S, Marincola FM. Lost in Translation: Obstacles to Translational Medicine. *J Transl Med.* 2004; 2:14. [PubMed: 15149545]
3. Kaitin KI, DiMasi JA. Pharmaceutical innovation in the 21st century: new drug approvals in the first decade, 2000–2009. *Clin Pharmacol Ther.* 2011; 89:183–8. [PubMed: 21191382]
4. Dudley JT, Schadt E, Sirota M, Butte AJ, Ashley E. Drug discovery in a multidimensional world: systems, patterns, and networks. *J Cardiovasc Transl Res.* 2010; 3:438–47. [PubMed: 20677029]
5. Lechler RI, Sykes M, Thomson AW, Turka LA. Organ transplantation--how much of the promise has been realized? *Nat Med.* 2005; 11:605–13. [PubMed: 15937473]
6. Khatri P, Sarwal MM. Using gene arrays in diagnosis of rejection. *Curr Opin Organ Transplant.* 2009; 14:34–9. [PubMed: 19337144]
7. Khatri, P.; Sarwal, MM. Functional Pathway Analysis for Understanding Immunologic Signature of Rejection: Current Approaches and Outstanding Challenges. In: Marincola, FM.; Wang, E., editors. *Immunologic Signatures of Rejection.* Springer; New York: 2010.
8. Sarwal M, et al. Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling. *N Engl J Med.* 2003; 349:125–38. [PubMed: 12853585]

9. Mueller TF, et al. Microarray analysis of rejection in human kidney transplants using pathogenesis-based transcript sets. *Am J Transplant.* 2007; 7:2712–22. [PubMed: 17941957]
10. Pham MX, et al. Gene-expression profiling for rejection surveillance after cardiac transplantation. *N Engl J Med.* 2010; 362:1890–900. [PubMed: 20413602]
11. Sutherland SM, et al. Protein microarrays identify antibodies to protein kinase Czeta that are associated with a greater risk of allograft loss in pediatric renal transplant recipients. *Kidney Int.* 2009; 76:1277–83. [PubMed: 19812540]
12. Butte AJ, Sigdel TK, Wadia PP, Miklos DB, Sarwal MM. Protein microarrays discover angiotensinogen and PRKRIP1 as novel targets for autoantibodies in chronic renal disease. *Mol Cell Proteomics.* 2010
13. Sigdel TK, et al. Shotgun proteomics identifies proteins specific for acute renal transplant rejection. *Proteomics Clin Appl.* 2010; 4:32–47. [PubMed: 20543976]
14. Ling XB, et al. Integrative urinary peptidomics in renal transplantation identifies biomarkers for acute rejection. *J Am Soc Nephrol.* 2010; 21:646–53. [PubMed: 20150539]
15. Sigdel TK, Sarwal MM. The proteogenomic path towards biomarker discovery. *Pediatr Transplant.* 2008; 12:737–47. [PubMed: 18764911]
16. Domanski L, et al. Hypoxanthine as a graft ischemia marker stimulates catalase activity in the renal vein during reperfusion in humans. *Transplant Proc.* 2006; 38:35–8. [PubMed: 16504657]
17. Vollenbroeker B, Koch JH, Fobker M, Suwelack B, Hohage H, Muller U. Determination of cyclosporine and its metabolites in blood via HPLC-MS and correlation to clinically important parameters. *Transplant Proc.* 2005; 37:1741–4. [PubMed: 15919451]
18. Annesley TM, Clayton LT. Quantification of mycophenolic acid and glucuronide metabolite in human serum by HPLC-tandem mass spectrometry. *Clin Chem.* 2005; 51:872–7. [PubMed: 15746301]
19. Wishart DS. Metabolomics in monitoring kidney transplants. *Curr Opin Nephrol Hypertens.* 2006; 15:637–42. [PubMed: 17053480]
20. Anglicheau D, et al. MicroRNA expression profiles predictive of human renal allograft status. *P Natl Acad Sci USA.* 2009; 106:5330–5.
21. Shi L, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol.* 2006; 24:1151–61. [PubMed: 16964229]
22. Morgan AA, Khatri P, Jones RH, Sarwal MM, Butte AJ. Comparison of multiplex meta analysis techniques for understanding the acute rejection of solid organ transplants. *BMC Bioinformatics.* 2010; 11 (Suppl 9):S6. [PubMed: 21044364]
23. Chen R, et al. Differentially expressed RNA from public microarray data identifies serum protein biomarkers for cross-organ transplant rejection and other conditions. *PLoS Comput Biol.* 2010; 6
24. Wang E, Worschech A, Marincola FM. The immunologic constant of rejection. *Trends Immunol.* 2008; 29:256–62. [PubMed: 18457994]
25. Li L, et al. Identifying compartment-specific non-HLA targets after renal transplantation by integrating transcriptome and “antibodyome” measures. *Proc Natl Acad Sci U S A.* 2009; 106:4148–53. [PubMed: 19251643]
26. Pascual M, Theruvath T, Kawai T, Tolkoff-Rubin N, Cosimi AB. Strategies to improve long-term outcomes after renal transplantation. *N Engl J Med.* 2002; 346:580–90. [PubMed: 11856798]
27. Suntharalingam G, et al. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N Engl J Med.* 2006; 355:1018–28. [PubMed: 16908486]
28. Puellmann K, Beham AW, Kaminski WE. Cytokine storm and an anti-CD28 monoclonal antibody. *N Engl J Med.* 2006; 355:2592–3. author reply 3–4. [PubMed: 17167145]
29. Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. *Nat Biotechnol.* 2004; 22:1253–9. [PubMed: 15470465]
30. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS computational biology.* 2010; 6:e1000662. [PubMed: 20140234]

31. Yang X, et al. Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat Genet.* 2009; 41:415–23. [PubMed: 19270708]
32. Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci U S A.* 2000; 97:12182–6. [PubMed: 11027309]
33. Barabasi AL, Albert R. Emergence of scaling in random networks. *Science.* 1999; 286:509–12. [PubMed: 10521342]
34. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature.* 2001; 411:41–2. [PubMed: 11333967]
35. Taylor IW, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol.* 2009; 27:199–204. [PubMed: 19182785]
36. Ashley EA, et al. Network analysis of human instent restenosis. *Circulation.* 2006; 114:2644–54. [PubMed: 17145989]
37. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature.* 2009; 461:218–23. [PubMed: 19741703]
38. Merali Z. Computational science:...Error. *Nature.* 2010; 467:775–7. [PubMed: 20944712]
39. Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol.* 2009; 5:290. [PubMed: 19638971]



**Figure 1.** Applications of translational bioinformatics in transplantation. The asterisk denotes the areas in transplantation in which translational bioinformatics is poised to play a crucial role. Image of the kidney adapted from, [https://secure.wikimedia.org/wikipedia/en/wiki/File:KidneyStructures\\_PioM.svg#metadata](https://secure.wikimedia.org/wikipedia/en/wiki/File:KidneyStructures_PioM.svg#metadata) available under Wikimedia commons license, (<https://secure.wikimedia.org/wikipedia/commons/wiki/Commons:Licensing>).