Pittsburgh Workshop on Quantitative Systems Pharmacology (QSP) in Personalized Medicine (PM)

Pittsburgh, PA on November 19 and 20, 2013

Co-Sponsored by the University of Pittsburgh, University of Pittsburgh Medical Center (UPMC), and Carnegie Mellon University

Supported by the University of Pittsburgh Institute for Personalized Medicine (IPM), the University of Pittsburgh Drug Discovery Institute (UPDDI) and the Clinical and Translational Science Institute (CTSI)

Jeremy Berg (Pitt) opened the Workshop with an overview of the meaning and goal of "personalized medicine", which has gone from personal physician-patient interactions to the harvesting of genomic and other patient data for both prevention and treatment of disease. Although the cost of sequencing has dropped dramatically, the cost of analysis remains high, the certainty in sequence calls remains limited, and our ability to interpret accurate variations is rudimentary. He laid out the partnership forged between the University of Pittsburgh and the University of Pittsburgh Medical Center (UPMC) to create an Institute of Personalized Medicine focused on the interpretation of variants. The preliminary initiatives will target pharmacogenomics, complex disease, cancer, and definitive and prenatal diagnosis.

Lans Taylor (Pitt) then moved to address QSP as a means to provide an integrated systems-level approach to determining the mechanism of new and existing drugs and their therapeutic application for personalized medicine. He emphasized that QSP addresses the complexity inherent in humans through the use of multiscale biased and unbiased approaches. He walked through the major components of the NIH White Paper on QSP, including the horizontal and vertical integration of systems biology and pharmacology within a QSP approach. He then introduced how the UPDDI was implementing QSP for drug discovery for personalized medicine. Inputs include patient samples and the data generated from them, leading to inferred pathways and ultimately an experimental model of the disease phenotype. Discovery activities include the use of phenotypic discovery using High Content Screening (HCS) platforms applied to the disease relevant, experimental models followed by an iterative process of profiling with HCS, application of computational methods to extract complex data, mathematically modeling the data, including the development of computational models of pathways involved in disease progression, formulating hypotheses and testing the hypotheses by HCS. Outputs include functional knowledge gained along the way, improved drug efficacy and safety, and more precise drugs with companion diagnostics. He then briefly noted some of the initial programs applying QSP at Pitt that harness the potential of phenotypic discovery using HCS, including an integrated programs pursuing biased (ER stress, protein misfolding, mitochondrial dysfunction) and unbiased approaches to determining causal pathways in three distinct disorders: metastatic breast cancer, non-alcoholic fatty liver disease, and Huntington's disease. He also briefly described

the development of a human, 3D liver acinus module designed to drive data generation for computational modeling.

Dr. Taylor concluded by laying out the key goals of the Workshop:

1. To characterize QSP and PM for the purposes of ongoing discussions

2. To understand the perspectives from academia, industry, and government on the main bottlenecks for the successful implementation of QSP and PM

3. To define what type of academic, industry, and government collaborations/partnerships can optimize the implementation of QSP and PM

4. To define what financial models will permit QSP and PM to be commercially viable

Following this introduction, Brian Shoichet (UCSF/UToronto) presented an elegant overview of classical pharmacology (when receptors were named after the small molecules acting on them) and its strengths (penetrance and systemic efficacy) and the progression to molecular pharmacology with a molecular toolkit and the addition of phenotypic discovery using model screening systems (zebrafish, C. elegans, cell lines/engineered model tissues) and Similarity Ensemble Approach (SEA), which quantifies classical pharmacology (choose query, get ligand set, compare with other ligand sets, rank by significance).

He walked through an example in which SEA was used to get 3000 targets for 25 fat-actives obtained through a phenotypic screen for fat build-up in C. elegans and then examine the ligands for the 3000 identified targets. Upon testing 20 predictions in vitro and genetically, SEA correctly predicted the ligand and mechanism of activity on the target receptor for 3 compounds (Lemieux, Keiser, et al. PLoS 2014 in press). As part of this same work, SEA identified 12 human targets for 11 fat-active compounds screened in C. elegans.

Here Dr. Shoichet moved on to address the barriers to further progress in QSP, namely: target-oriented libraries and orthogonality, the chemical basis of pharmacology, and biorelevant novelty. For the first barrier, he pointed out that target-oriented libraries exist at pharmaceutical companies and expressed his desire that the means for sharing them with academic researchers be found. During the discussion on Day 2, Jim Stevens (Eli Lilly) pointed out that Eli Lilly established compound-sharing agreements with multiple research institutions, where one challenge was giving the company 30-day advanced review of manuscripts to be submitted for publication. Indeed, the barrier of sharing between industry and academia arose throughout the workshop and is summarized at the end in terms of the need to align incentives. Outside industry, progress is being made as well, such as the Cancer Therapeutics Response Portal (http://www.broadinstitute.org/ctrp/).

Dr. Shoichet also emphasized the need to establish orthology between model systems and humans beyond gene sets, to encompass signaling pathways themselves, such that circuit orthology supports greater confidence in targets being shared between the model system and humans as well.

In addressing his second barrier, he summarized new work demonstrating that the activity of chemicals on biological circuits cannot be assumed based on similarity. He illustrated the mapping of GPCRs both by sequence ID and by ligand similarity, which were quite different and resulted in 6 predicted crosses that were tested experimentally and confirmed. Their analysis beautifully illustrates the polypharmacy of synthetic molecules and their wide-ranging promiscuity in the same receptor class, which he went on to explain is not an accident. GM Tompkins demonstrated an ancient code in which signaling receptors evolved around their messengers, such that the same metabolically expensive small molecule (e.g., serotonin) can signal in multiple receptors in different time domains, depending on the task, to accommodate various needs in the cell. Such an evolutionary process highlights the need to glean biorelevant novelty from target-oriented libraries.

As part of this discussion, Dr. Stevens noted that polypharmacology, although a dirty word in some circles (Dr. Shoichet confirmed resistance to the magic shotgun approach of Roth et al.), must be accepted, but noted that the challenge to be addressed is how to advance a molecule as part of polypharmacology. He also noted that his take-home message from Dr. Shoichet's presentation is that we should be looking to humans as a source of natural products for identifying molecules, which Dr. Shoichet agreed would already be known to be biologically active and relevant. In other words, human-based, personalized medicines.

The discussion also touched on how to better computationally predict which compounds with similar structure and biophysical properties will be biologically active. Dr. Shoichet again noted that chemical information for molecules and even sets of small molecules is impoverished and that closer engagement is needed with experimentalists and experimental model systems. In other words, QSP.

Jim Stevens (Eli Lilly) started off with a review of the history of drug development science and regulation, reiterating the time and expense of identifying 24 molecules, of which half will be advanced to animal testing, with the goal of submitting one to the FDA. The goals in this process are to avoid unnecessary risks (minimize toxicity risk in selecting molecules), assess emerging risks (preclinical and clinical trials), and manage identified risks (launched). Overall, the emphasis is to reduce false negatives (agents we thought were safe and whose use we promoted). The current paradigm of safety testing, in which 95% of patients must be protected, overpredicts risk and removes promising leads from the pipeline (leads that might have otherwise been useful in a personalized medicine paradigm).

To address this barrier, he asked, if I can't know a lot about a molecule, what do I need to know? He started with an example for which we have the information – preclinical toxicity testing – in terms of toxicity versus tissue concentration and the combined impact on safety. He then moved on to transcript profiling methods (e.g., support vector machines, Connectivity Map, weighted gene co-expression network analysis) to reduce the dimensionality of the data and in turn the size of the probe set without losing information (i.e., know the mechanism is not random). For example, out of 31,000 probe sets, fewer than 10,000 might be active in liver tissue, and when these are analyzed as co-expressed gene sets, each module would include only those genes (e.g., ~18) that respond similarly to the drug through a known mechanistic pathway.

Dr. Stevens concluded with the question as to how QSP could be used to make better decisions in industry. His suggested responses were simple and straightforward:

- Pick the right question using a data driven approach
  - Data + QSP + Knowledge = Transformation
  - Collect new data and measures focused on outcomes
  - Be quantitative

- Focus on short term and long term objectives – the patient is waiting
  - Drive toward applications in the short term
  - Don't use a complex tool to answer a simple question
  - Write requirements for new approaches while using what we have
  - Enable transformation in the longer term

- Design the right tools
  - Accessibility/Usability
  - Converge methods to the task – e.g. classifiers vs network, supervised or unsupervised, noise filtering
  - Public vs proprietary

- Tailor the tools/tests to the important decisions
  - Avoid risk –vertical integration and predictions
  - Assess emerging risks – vertical/horizontal integration and mechanism of action
  - Manage existing risks - horizontal integration and mechanism of action

The discussion focused mainly on the right experimental tools for assessing drug activity. Dr. Stevens expressed his strong support for homogeneous animal models, particularly in assessing exposure (more difficult to computationally model than drug distribution), though he recognized the temporal constraints of assessing, for example, ER toxicity using individual snapshots from which the timeline of activity must be extrapolated. Dr. Taylor suggested human, 3-D organs on a chip as a means

of obtaining better insight into human response, which Dr. Stevens felt lost the complexity of cell-cell and systemic signaling. Although Dr. Berg commented on personalized rat medicine to emphasize growing awareness of differences in gene expression between species, Dr. Stevens still felt in vivo models were the faster, better understood approach

Mike Rogers (NIGMS), in introducing the NIH perspective on QSP, started off by touching on the previously discussed issues of multiple targets, polypharmacology, and chronopharmacy (timing of the body's reaction to drugs) and noting that NIGMS recognized the need to take action as the number of new drugs coming on the market dropped. The Institute's response was to address the underlying complexity by merging its pharmacology and systems biology portfolios with those in mathematical modeling.

He reviewed the history of the two NIH workshops on Quantitative and Systems Pharmacology in 2008 and 2010 and the preparation of the White Paper in 2011. He noted especially the focus on including diverse attendees, especially those in pharmacokinetics and pharmacodynamics, who were already modeling at the human level and drilling down to mechanism. He summarized the components of the definition of QSP:

- An emerging discipline focused on identifying and validating drug targets
- Understand how drugs act in space and time
- Understand and predict therapeutic and toxic effects of drugs
- Formal mathematical and computational models, multi-scale models
- Has an intrinsic and extensive experimental component
- Interactions among multiple elements (biomolecules, cells, tissues, etc.)
- Draws on several existing disciplines and incorporates approaches from tissue and organ physiology, pharmacology, and cell biology as well as newer areas of bioinformatics and various omics

He also reviewed the Working Group's recommendations for research and training in QSP as well as progress made toward these. At the university level, these include an Initiative in Systems Pharmacology at Harvard and a Center in Quantitative Pharmacology at UCSF. The American Society for Pharmacology and Experimental Therapeutics has added a journal and a pre-meeting workshop on systems pharmacology, and the European Federation for Pharmaceutical Sciences held a workshop on the formation of a QSP research network. At the NIH, NIAID, NCI, NIA, and the NIH-wide Biomedical Information Science and Technology Initiative (BISTI) have all embraced QSP projects. NIGMS itself expanded its Centers for Systems Biology (P50) program to include QSP (as embraced by the Centers at Mount Sinai and UCSD), will transition its support of the Pharmacogenomics Research Network to different funding mechanisms (R24, U01), and will launch a new Centers for Personalized Medicine (P50) initiative. He also reviewed NIH initiatives on Big Data

to Knowledge (BD2K) and the Druggable Genome, which focuses on un-annotated proteins from 4 highly targeted gene families (GPCRs, kinases, ion channels, nuclear receptors). He anticipates both more funding opportunities and greater representation in research portfolios for QSP, in particular with multi-PI applications and industry-academia partnerships.

Trey Ideker (UCSD) led into his presentation on molecular networks by first introducing the Center for Bioinformatics Analysis within the Institute for Genomic Medicine at UCSD, with its Executive Director fulfilling the role of a VP at a drug company, more focused on products and services than pure research, as well as his own group's software Cytoscape, an open source Java platform for the integration and interpretation of omics data around networks and pathways.

The need for software development in such analyses is evident through the discovery that the average solid tumor has 40-80 somatic mutations, only a subset of which (probably fewer than 10) lead to cancer, and that no two tumor genotypes (even within the same patient) are alike. For example, an analysis of ovarian tumor genomes in TCGA found that while ever tumor had a TP53 mutation in chromosome 17, there was otherwise no pattern to the remaining mutations. Thus, it is critical to stratify tumor genomes using networks and tools such as Pathway Commons (MSKCC), HumanNet (UT Austin), and StringDB (EMBL); that is, look at the effect of mutations on neighboring pathways, not just the pathway involving the mutated gene, and cluster by how close a pathway is within the network to the mutated gene.

Dr. Ideker walked through the example of ovarian tumor subtypes stratified by aggressiveness (survival) and cisplatin resistance. In looking at the aggressiveness subtype network, almost none of the tumors have the same mutation, but mutations are clustered near each other in the overall network, with the rest of the network relatively unaffected. He pointed out that while Titan, the longest gene, was often hit, so were other proteins near Titan in the network. He used this example to emphasize in his summary the need for network knowledge and the hierarchy of modular components (i.e., an ontology). He views an interesting emerging problem as how to assemble such ontologies from omics data. In considering these networks in the context of the tumor microenvironment, he felt the additional of expression profiles would be important as well as the use of ENCODE data to examine regulation of gene expression.

Gary Nabel (Sanofi) shifted the focus from omics to immune repertoires and the expanding universe of antibody therapies, including the growing number and diversity of disorders treated with monoclonal antibodies (e.g., rheumatoid arthritis, multiple sclerosis, breast and colorectal cancer, Crohn's disease, postmenopausal osteoporosis, psoriasis, etc.), which was a $45 billion global market in 2011.

Dr. Nabel jumped straight into an example from his former lab at NIAID to illustrate how they defined the CD4 binding site, for which initial contact is on the inner portion of the outer domain of the HIV-1 envelope and extends down to the bridging sheet. Their group resurfaced stabilized glycoprotein cores specific for the conserved CD4 binding site (altered other surface residues to eliminate reactivity with non-neutralizing Abs) and used these probes to identify B cells from an HIV-1-infected patient that expressed neutralizing Abs specific to the CD4 binding site. They identified 3 monoclonal Abs, including VRC01, which neutralized over 90% of natural circulating HIV-1 virus. They determined that VRC01 worked so well through its partial mimicry of CD4 binding to glycoprotein 120, including the conformationally conserved bridging sheet attachment. In experimental confirmation of their neutralizing activity, VRCO1 protected non-human primates against both rectal and vaginal challenge with SHIV162P3. His group then used VRC01 as a model to understand how broadly neutralizing Abs develop naturally and discovered a common evolutionary tree in multiple HIV-1-infected patients. The maturation of progressive intermediates leading to VRC01 suggests a novel strategy for guiding the host immune system to generate its own VRC01 through intermediate mutations and thus eliminate off-target hits.

Dr. Nabel then moved on to another public health priority in which a universal immunogen would be invaluable: influenza vaccine, which currently requires 120-150 million seasonal doses each year at a cost of $2.8-4.0 billion. As with their focus on the structure of the CD4 binding site, his group examined the sites of broadly neutralizing Ab binding in the stem of glycosylated influenza hemagglutinin and identified the potential for spike-like structures to take the place of the Abs. They fused viral hemagglutinin to ferritin nanoparticles, which self-assembled to create trimeric spikes that fit into the vulnerable portions of the viral stem normally targeted by broadly neutralizing Abs. However, this ferritin nanoparticle platform neutralized H1N1 viruses from 1934 to 2007 and protected ferrets from an unmatched 2007 H1N1 virus challenge, pointing to its promise as a universal synthetic vaccine as well as the promise of harnessing structural biology, genomics, and patient selection (to study the evolution of Ab intermediates) for future vaccine development. With regard to the potential for genetic drift, he noted glycan differences in avian versus human viruses but agreed that preemptive protection will need to take into account the prediction of genetic drift.

Ziv Bar-Joseph (CMU) moved from mining structural data for leads to mining and integrating large omic databases for QSP. Continuing on with the infectious agent theme, he presented 3 examples with wide-ranging effects in mice, non-human primates, and humans in asking why we can tolerate some infections but not others: Mycobacterium tuberculosis, influenza virus, Francisella tularensis. In discussing the use of omic data from 3 species, he noted that expression measurements are noisy, ignore interactions and networks, and provide no temporal data, which makes direct comparisons to identify a common signature problematic. Instead, his group has developed a method to first combine the omic data from 3 species

injected with 2 bacterial and 2 viral pathogens and then deconvolute them into pathogen-specific and species-specific dynamic networks (as protein-DNA interaction, binding motif, and gene expression data) for understanding innate immune response. He showed differences in the timing and direction of activation of the transcription factor IRF7 as one example, as well as applications of his approach to examine stress and hormone response, stem cell differentiation, and fly development.

Dr. Bar-Joseph then moved on to another project with Dr. Ideker involving the integration and analysis of gene expression patterns over time from 3 evolutionarily distant species of fungus (Saccharomyces cerevisiae, Candida glabrata and Kluyveromyces lactis) following exposure to the antifungal agent fluconazole. Their approach used a novel soft clustering algorithm that concurrently clustered data from all species while incorporating sequence orthology. In comparing the dynamic transcriptional responses among the 3 species, they identified significant divergence in regulatory programs for transporter usage by fluconazole-resistant strains versus the susceptible strain.

He next demonstrated the use of the search engine through a narrative about mice, aging, and the feminization effect - that is, their analysis showing that the gene SIRT6 regulates lifespan in male mice only. First, he backed up to ask if we can use other people's expression data for such analyses, given that nearly one million gene-expression data sets are now publically available. To do so, his group developed Expression Blast, which can mine data from more than a million arrays by downloading data from Gene Expression Omnibus, identifying gene IDs, scaling the values, merging technical replicates, and identifying and matching treatment versus control cases. This identical processing of all the data allows comparisons across experiments and species.

Getting back to the regulation of age in male mice, Dr. Bar-Joseph's group used ExpressionBlast data of male mice overexpressing Sirt6. Although phenotypic effects of Sirt6 have been documented (e.g., increased lifespan and reduced levels of diet-induced obesity), little is known about the specific mechanisms by which it operates or the pathways it regulates. ExpressionBlast allowed them to discover other mouse experiments whose results had significant similarity to or were negatively correlated with the Sirt6 overexpression profile and thus help determine the functional categories and pathways activated by Sirt6 (Icam-1, iNOS, IL-6, IL1b, CD38).

Although ExpressionBlast allows users to mine large, unstructured expression databases, Dr. Ideker wondered how to address comparisons of heterogeneous data from different chip manufactures and sequencing technology, different timepoints, different species, and so on. In addition, John Wikswo (Vanderbilt) wondered about how metabolomics could be brought into the analysis, to which Dr. Stevens noted the metabolome comes mostly from the microbiome, adding another layer of complexity. Dr. Bar-Joseph noted that the goal was to combine and deconvolute all

data to obtain predictions of cell regulation of metabolism but that mechanism would take additional analyses.

Ivet Bahar (Pitt) gave an overview of computational pharmacology in which network pharmacology is complemented by molecular pharmacology (e.g., siRNA, high-throughput screening, pharmacophore modeling). She began with a figure highlighting all known FDA-approved drug-protein target associations and their interrelationships. Indeed, many databases catalog drug-target or protein-ligand associations: DrugBank (6,711 drugs, 4,227 targets), ChEMBL (1.2 million compounds, 9,570 targets, 3 million bioassay measurements from 40,000 publications), STITCH (300,000 chemicals, 2.6 million proteins). She then noted the use of similarity ensemble approach (SEA) to cluster proteins based on the similarities of their ligands and thus identify new targets for existing drugs.

However, the level of integration versus modularity of proteins and the combination of proteins and their modular domains gives rise to the tremendous diversity and complexity in phenotype. Further, proteins use conformational transitions (i.e., allostery) to alter their signaling and function; for example, two different conformations of GPCR bind to two different agonists that branch into two different pathways. Allostery can thus both cause disease (cellular level) and contribute to drug discovery (molecular level). The multiplicity of pathways involved and resultant phenotypes expands depending on the targeted surface region (accessible structural change/dynamics) as well as interactions with different upstream or downstream substrates. Another factor in drug-target interactions relates to drug promiscuity, which can be beneficial if the other targets are o-in the same pathway or complimentary pathways, but can also be potentially harmful.

Dr. Bahar presented her lab's approach to addressing these issues through active learning by probabilistic matrix factorization (PMF). They deduced latent vectors characteristic of each drug or protein through a PMF analysis of the complete dataset of drug-target interactions. Using 1,576 targets, 1,509 drugs, and 5,630 interactions in DrugBank, they visualized clusters of drugs and targets based on their overall interaction profiles – that is, their phenotype. Clustered drugs exhibited similar therapeutic actions despite very dissimilar structures. The deduced latent variables capture similarity in therapeutic action between different drug clusters. In looking at drugs that target enzymes, ion channels, GPCRs, and nuclear receptors, their PMF predictions outperformed existing published methods for the same analyses with an overall accuracy of 86% compared with 78-83%. When they applied PMF to DrugBank for predictions on drug repurposing, 89% of the top 100 predictions are accurate, formerly hidden, interactions. Having trained their model with a subset of interactions to achieve 89% accuracy, they used the full dataset to generate new de novo predictions for drug-target interactions and confirmed 7 of these experimentally.

In collaboration with David Perlmutter, they sought to identify drugs that might be repurposable to prevent alpha 1-antitrypsin-Z (ATZ) accumulation in AT deficiency (ATD). They trained a logistic regression classifier to identify genes associated with ATZ aggregation, found matches with 44 of 54 genes checked in WormBase, and retrieved sequences of known drug targets from DrugBank. Using BLASTp, they aligned the sequences to identify 3 ATZ targets with high sequence similarity to known drug targets and went back to DrugBank to get drugs that inhibited the top three ATZ targets (MRP1, MRP3, BSEP). They next found human orthologs and used RNAi screens to validate activity for each of the interacting drugs tested.

Having explained the framework for her approach to discover new targets for existing drugs, Dr. Bahar noted that future work will need to address the network of interactions and checking overlap among multiple pathways to ensure that hitting one target does not negatively perturb an unintended pathway. Her group is developing application program interfaces (APIs) for this and to examine polypharmacological strategies. She and Dr. Shoichet specifically asked whether pharmaceutical companies were interested in drug repurposing as well as the legal and financial ramifications of repurposing. Dr. Stevens noted that repurposing is of interest, especially to rescue abandoned molecules, but that the legal issues were greater than the technical issues, especially with regard to patent life (too short to support return on investment). This would also be true for low-dose polypharmacology to reduce side effects (as type of drug repurposing).

Steve Shapiro (UPMC) then moved the focus from in silico drug discovery to in silico patient management, starting with his futuristic representation of a typical internist day in 2018 and the historical and immediate data (all types) available on his tablet in assessing and developing a care plan for a 57-year-old male smoker with a lung nodule. Getting to this point must get past the cost barrier, since the US cannot sustain rising costs in health care, while quality suffers from fragmented care and poor health habits at the population level. One benefit he foresees is that technology will be able to reassure patients that less can be more (i.e., patient feels confident that they are receiving standard of care even if physician does not prescribe drugs, order procedures, etc.).

Dr. Shapiro then moved on to review the UPMC model as a hybrid organization (structured for business in multiple sectors: heathcare provider, payer, researcher-innovator) as one well positioned to move forward in the era of healthcare reform. The size (60,000 employees, >$11 billion operating budget), scope (22 hospitals with 4500 beds, >400 outpatient sites), volume (260,000 admissions, >5 million outpatient visits, 2.2 million members insured), and reach (18 countries, multiple corporate and academic partners) allows the UPMC model to coordinate care, align incentives, and commercialize innovation in one vibrant ecosystem.

The UPMC strategy for embracing personalized medicine combines smart technology for managing and exploiting big data with good science and physician-

scientists. Through its close partnership with Pitt, UPMC can rapidly infuse proven research and protocols effectively treat disease while reducing over-diagnosis and over-treatment into their current practice of medicine. At the same time, given its size and scope, UPMC can focus on meeting population needs and serve as a public health laboratory for reducing costs and enhancing value. Because UPMC is both a provider and a payer, they continuously address both sides of the equation to deliver better patient outcomes at lower costs.

In creating a patient-centered medical home, UPMC envisions a team that provides coordinated comprehensive care and re-engineers team member responsibilities to address preventive, acute, and chronic care through one team. This continuum of care will also be continuous through enhanced communication and expansion of telemedicine, which will be possible by aligning incentives to reward non-procedural care with improved quality and outcome, and this will be possible by capturing electronic data for analysis.

Indeed, recognizing the need to address steadily escalating volumes and diverse types of data and the need for real-time analytics, UPMC is in year 2 of 5 of an Enterprise Analytics initiative to ingest over 200 data sources, aggregate and harmonize the data (both structured and unstructured), and support advanced data mining (such as that discussed by Dr. Bar-Joseph). Big data analytics will allow clinical care to be redesigned to reduce variation by ensuring standardized care using clinical pathway protocols backed by data and holding accountable those physicians who do not follow these clinical pathways (though they are constructed to permit variation when the evidence is not as strong). These clinical pathways are optimized first for quality (outcome, patient satisfaction) and then cost and refined as appropriate to reduce cost while maintaining the highest quality.

Dr. Shapiro walked through an elegant example in making this case, and particularly the point of studying data within the UPMC ecosystem versus relying solely on recommendations in the published literature. In developing the clinical pathway for the cardiac catheter lab, they examined the use of an aspiration catheter (Angiojet) to remove clots from an occluded artery prior to stenting. In comparing individual cardiologists, they discovered that academic physicians were more likely than full-time clinicians to perform aspirational thrombectomy. Although the additional supplies required for the aspiration catheter raised the per-case procedure cost by $897, this was offset by savings in the form of reduced length of stay (1.9 days shorter) in patients in whom the Angiojet was used (~$855/case). Both mortality (10.2% vs 15.1%) and need for repeat revascularization (5.5% vs 14.0%) were lower at 6 months in patients in whom aspirational thrombectomy was performed. Given that repeat revascularization within 6 months was performed in 2% of patients with similar length of stay and supply costs as the initial procedure, the savings based on reduction in readmissions supports the use of Angiojet at UPMC. However, a prospective clinical trial in 7000 patients published in NEJM (2013;369:1587) reported no significant 30-day mortality benefit but a 50% decrease in re-stenosis, and the conclusions in the accompanying editorial focused

on the lack of benefit in saving lives, suggesting that catheter thrombectomy "remains an unmet aspiration." If UPMC had not done its own analysis of its routine clinical data, it might have missed an opportunity to reduce cost, morbidity, and mortality. Although such enterprise analytics are powerful tools in improving care, they will not replace clinical trials.

Dr. Shapiro concluded with an overview of UPMC's strategy to move into personalized medicine, starting with its support of the University of Pittsburgh Institute of Personalized Medicine, which will initially focus on initiatives in cancer, complex disease, pharmacogenomics, and prenatal and neonatal care. The goal is to collect both phenotype and genotype data for all patients to identify the most appropriate and effective care for each. Issues to be addressed include how much and which data to collect and tests to run, how to handle informed consent, what data to store and how, and funding or reimbursement models for all aspects of the process. One target for cost savings will be the development of biomarkers that help avoid over-diagnosis and care capable of changing the patient perspective that more is better.

Andrew Plump (Sanofi) reminded us again of the cost of drug development (5.125M Euros per approved drug), with almost half (47%) of this being spent on target identification and validation, lead identification, and lead optimization and a third (33%) spent on preclinical and early clinical testing. He used cardiovascular disease as a success story with two sides: mortality due to cardiovascular disease has declined significantly in the last 30 years due to medical intervention, but the global incidence is expected to double by 2050 due to the Westernization of lifestyle in Asia especially. He noted several recent high-profile failures in late phase III testing (in trials costing $500-800 million), which he described as "lost in translation" due to the pharmaceutical industry's continued use of heterogeneous populations. He emphasized that this traditional approach must be replaced by a patient-centric approach to treat actual risk rather than high risk factor levels and to design and implement mechanism-based treatment in patient segments selected based on their underlying disease mechanism. Achieving this goal will require developing better markers of risk and the right targets.

Pharmaceutical companies start by looking at an unmet need, such as the fact that 30% of patients with elevated LDL levels have not reached their therapeutic goal. This observation points to the need to understand the disease in these patients to choose better targets and develop appropriate predictive biomarkers. Thinking back to the aspirated thrombotic plaque that Dr. Shapiro just discussed, Dr. Plump suggested that this plaque be used like a tumor biopsy to mine for potential biomarkers and in particular better predictors of outcome than imaging, blood pressure, obesity, and lipid profile. He presented data showing that gene expression profiling of carotid atherectomy samples could differentiate between non-diabetic and diabetic patients, with the latter cohort showing significant enrichment of up-regulated genes responsible for immune stimulation. They in fact section the plaque

into 1 mm slices, with one used for histology, the next ground up to analyze the transcriptome, and the third used to identify soluble biomarkers (repeated for each cluster of 3 adjacent sections). Thus far, they have found wide variability across the length of the plaque, with markers related to reactive oxygen species, inflammation, endothelial dysfunction, extracellular matrix production, and other cell processes critical to thrombotic plaque development and host response. As a specific example, he reported that Cathepsin B correlates with plaque macrophage and total cholesterol, suggesting its value as a potential molecular imaging marker as an indicator of a plaque that is about to rupture.

He then moved on to oncology, where a critically important decision is the choice of target. Pharmaceutical companies traditionally select targets based on what they have experience with, then figure out the drug for that target – and then the disease in which to apply the drug, with little concern for patient characteristics beyond the disease. However, the blockbuster failures he described at the start of his talk demonstrate the limited efficacy of this approach and the need for a more translational approach. This revised model of drug development starts with the disease and patient cohorts to identify important targets and moves to understand the underlying mechanism so the drug can be developed precisely for that mechanism and that category of patients. He again illustrated this with an example.

Loss of p53 function is a universal characteristic of all cancer. Such inactivation can occur through genetic mutation or, in a small percentage of cases, due to amplification of HDM2, an E3 ubiquitin ligase that is massively overexpressed in tumors and has a binding pocket for p53. Such HDM2 amplification occurs in over 90% of patients with de-differentiated liposarcoma (DD-LS), providing a target with a target (binding pocket) in a specific disease and patient population. Sanofi is developing SAR405838, a potent HDM2 ligand that blocks binding to p53, so p53 remains active as tumor suppressor. Preclinical studies have confirmed the ability of SAR405838 to cause regression in DD-LS. Through this example, Dr. Plump noted that QSP was essential for target identification and early development.

At this point, Dr. Ideker wondered how big pharma was still in business, to which Dr. Plump responded that they are simply buying each other to buy leads, though doing so only buys so much time. Payer pressures on the industry to control costs are growing at the same time research costs are going up. Companies need to get better at observing the signals that an agent is failing earlier in the process; these signals are often identified retrospectively but were not recognized at the time. There was also briefly discussion of the importance of moving tumor cells straight from humans to mice, since culturing in plastic dishes is increasingly recognized as a source of new mutations.

Adrian Lee (Pitt) opened his discussion of data integration and sharing with a slide illustrating the complexity that must be embraced to take full advantage of the data available, which can only be done through QSP. He used NCBI milestones to

illustrate how genomic data and resources (base pairs and daily users) have grown exponentially over the past 23 years, and illustrated the growth in clinical record data by graphing the 1,003% increase in data storage at UPMC over the past 6 years (at 4500 Tb in 2013 – before the addition of genomic data). He gave an overview of the personalized medicine ecosystem created by UPMC to integrate and harmonize payer, provider, and research data so they can be searched and extracted efficiently. This system has already been shown to work for the integration of clinical and genomics data from The Cancer Genome Atlas (TCGA).

All these data need to be shared, and, indeed, those who share their omics data are rewarded with significantly higher citation rates. Dr. Lee discussed new methods for conducting research through shared data, such as crowd-sourcing competitions (prizes not needed – "players" take on the challenge for reasons of altruism, participation, and personal challenge/satisfaction) and gaming.

He again moved to the local UPMC model for managing and sharing data, which makes possible analyses of tumor and metastases genomes over time (from initial diagnostic biopsy through surgeries and subsequent biopsies to autopsy) and the discovery that the metastasis genome differs from that of the primary tumor.

The Pittsburgh Genome Resource Repository (PGRR) allows more efficient use of TCGA because the entire dataset (577 Tb) is stored locally and analyzed through the Pittsburgh Supercomputing Center and the Pitt Center for Simulation and Modeling. A PGRR IRB protocol permits linkage of TCGA data with UPMC patient data for those patients whose tumors have been submitted to TCGA (Pitt/UPMC is the largest contributor to the TCGA). Multiple data use agreements were also negotiated to make this system possible, and new software and graphical user interfaces are being developed to fully exploit all these data for research and clinical care. The framework for the PGRR data use agreements will in turn help support data sharing on a larger scale through the Western Pennsylvania Health Information Exchange, which has a goal to participate in the Nationwide Health Information Network.

Dr. Lee concluded by highlighting applications of genomic-guided clinical decision support for both research and clinical care. Patients with breast cancer at UPMC benefit from the use of the Ion AmpliSeq™ Cancer Panel, which targets 50 genes and provides details on the mutation, exon, protein, cDNA, mechanism, and potential drug for exploiting the mutation in a 10-hour process (construct library, prepare template, run sequence, analyze data, annotate results). From the literature, he gave an example from MD Anderson in which patients with diverse tumor types and a median of 5 prior therapies for whom a new therapy could be matched with their tumor molecular aberration(s) experienced a higher overall response rate, longer time-to-treatment failure, and longer survival. He noted that sharing data and experiences with genomic-guided medicine will be easier with resources such as openSNP, Patients Like Me, and Genetic Alliance and that the next-gen Tumor Board will be "molecular", with more computers and fewer white coats. New systems to

make this happen will be needed for storing, transferring, and analyzing data and integrating and analyzing genotype and phenotype data.

Ravi Iyengar (Mount Sinai) began with the observation that no speaker thus far had addressed the topic of drug activity itself – and proceeded to do so. He started by contrasting the classic view of drug action (one drug, one target, with all other interactions considered "off target" with separate pathways leading to either therapeutic or adverse effects) versus the systems pharmacology view of drug action (one drug, one primary target and multiple other targets with interacting pathways and both therapeutic and adverse effects). In support of the systems pharmacology view, he presented the single connected island of 481 nodes (GPCR drugs and targets) discovered in an a network analysis of all FDA-approved drugs and their targets; the most nodes in the remaining 178 islands discovered was 31, followed by one island with 23 nodes, and the rest having fewer than 20.

Dr. Iyengar then framed the challenges in systems pharmacology:

- Primary drug interaction is at the molecular level, but effects are at the tissue/organ or organismal levels.
- Co-relating molecular changes (genomic and epigenomic) to tissue/organ level effects has limited predictive power.
- We do not understand how meso-scale organization (cellular networks and tissue level networks) process information from molecular interactions to organismal functions (phenotypes).

His suggested approaches to addressing these challenges included network analyses to understand the topological context of drug action as well as multi-scale quantitative dynamical models of drug action.

For example, his group used the FDA Adverse Event Reporting System to determine whether, if Drug A is likely to evoke an adverse event, is there a Drug B that can mitigate this unwanted effect. Specifically, they looked at drugs with the potential to reduce rosiglitazone-associated myocardial infarction (MI) (with or without concurrent use of metformin, in patients with and without Type 2 diabetes) and discovered that exenatide significantly reduces the risk of this serious adverse event in all 3 categories. In analyzing the overlap between drug targets that reduce risk of rosiglitazone-associated MI, they discovered a clotting-related subnetwork that suggested the mechanism of action: namely, that exenatide reduces rosiglitazone-induced increases in PAI-1 levels, which they confirmed in diabetic mice (including the reduction in clot formation and firmness).

Based on this success, they examined the entire database for potentially beneficial polypharmacology, including that involving unrelated drugs, and discovered 19,133 combinations in which one drug could potentially reduce the frequency of adverse

events associated with another drug. They discovered, for example, that taking fentanyl with propofol significant reduces the risk of the anesthetic causing anaphylactic shock, that taking an H2 blocker significantly reduces the risk of SSRI use resulting in completed suicide, and that taking acetaminophen with niacin significantly reduces the risk of flushing. However, the value to drug companies of pursuing such discoveries is limited unless they also manufacture the mitigating agent, leaving this important analysis in need of alignment of incentives.

Dr. Iyengar then moved on to consider the human interactome, which encompasses all known direct protein-protein interactions of human gene products (and hence is continually expanding). His group focused on mapping the neighborhood of genes associated with long QT syndrome (LQTS) (ion channels and associated proteins) using mean first-passage time scoring to measure the functional distance between nodes. By combining their LQTS neighborhood with published GWAS data, they identified new polymorphisms likely to affect the QT interval. He also used this neighborhood to identify nodes that could result in an adverse cardiac event when targeted by specific drugs. They validated their model using human data by ranking their predictions by highest scoring target in the LQTS neighborhood and then screening for drugs that caused at least one QT event when used alone in the FDA Adverse Event Reporting System; were not listed in AZCERT QT drug lists; and did not list arrhythmic activity or toxicity in DrugBank. He highlighted the ability of this analysis to identify drugs used for different indications (e.g., antidiarrheal loperamide and anticancer dasatinib) with different possible mechanisms for lengthening QT interval.

While powerful, this network analysis is limited due to its inability to account for drug dosage, the impact of multiple genes on one drug, or the impact of epigenetic changes. To address this challenge, Dr. Iyengar's group has developed enhanced pharmacodynamic (ePD) models that integrate genomic and epigenomic data using multicompartment ordinary differential equation models of drug action. These ePD models take into account networks involved in both therapeutic and adverse effects as well as genomic changes that affect regulatory motifs.

He presented as an example an operational model of how an EGFR antagonist would affect tumor size when different pathway components are influenced by different genomic and epigenomic changes (e.g., copy number variant that increases EGFR activity, hypermethylation of *RASAL1* promoter region resulting in decreased RASAL1 concentration and increased RAS activity, and/or increase in microRNA 221 expression resulting in increased cyclin D-dependent CDK4/6 activity). He then used the ePD to model a cohort of cancer patients with the same initial driver mutation to show that although an EGFR kinase inhibitor might result in 80% inhibition of receptor activity across all patients, individual responses span from full remission to complete resistance, depending on the number and type of additional genomic and epigenomic variations. The take-home message was that looking at the drug target is just the starting point: complexity along the way affects the actual outcome.

Their ePD models provide a mechanistic understanding of why such variability occurs and can integrate a range of variations to predict drug response. These models contain representations of all pertinent regulatory networks and motifs and drug targets in the biological system of interest while also incorporating the effects of genomic, epigenomic, and posttranslational changes on model parameters (made possible because the outcome of all alterations are characterized as change in reaction rate or reactant concentration).

With the appropriate data, alterations at the various levels of biological regulation can be readily and precisely represented in ePD models. The data needed for ePD models include experimental data relating genomic, epigenomic, and transcriptomic changes to changes in protein levels and interactions within pathways and networks; a database of kinetic constants and concentrations of proteins in different cell types (quantitative human interactome); and measurements of different types of genomic changes for various diseases and drug treatment protocols (e.g., TCGA). In the future, he envisions two levels of predictive multiscale dynamical modeling: Level 1 to determine how drug interactions with targets affect cellular function in the context of multiple genomic and epigenomic changes, and Level 2 to connect cellular function to tissue- and organ-level function (including parameters that account for tissue organization and spatial characteristics).

Kathleen Bove (GE) moved the focus from therapeutics to diagnostics, providing an overview of GE's investment in imaging and other diagnostic equipment ($8.3B), information technology ($6.1B), and molecular medicine ($3.7B) with a goal of maintaining its leadership in precision diagnostics. She noted that the global market for personalized medicine and patient-based decision support was estimated to reach $810B by 2015. Indeed, she highlighted that the FDA will require companion diagnostics to approve new drugs (currently the case for cancer, to expand to other disorders) and gave as an example the simultaneous submission and FDA approval of Pfizer's XALKORI (crizotinib) with Abbott Molecular's ALK companion diagnostic.

Diagnostic data themselves will continue to increase in complexity, from a single sample-single result-single interpretation model to multiplexed and multiple omic data with multisource interpretation that will have high value but also high reimbursement and regulatory risk. GE is currently developing its M ultiOmyx™ platform to label up to 60 proteins on a single slide: the process first acquires background autofluoresence, applies 2-4 fluorescently labeled antibodies (Abs), then chemically inactivates these dyes, re-aquires a background fluorescence image, and applies a new set of 2-4 Abs labeled with the same 2-4 fluorescent dyes, repeating this cycle until all desired proteins have been labeled and imaged. Thus, 3 adenocarcinoma samples that appear similar with H&E staining demonstrate striking differences in AKT, cMET, ERK, and pERK in the MultiOmyx overlay images. Similarly, staining 44 proteins from 747 colorectal cancer patients revealed

extensive tumor heterogeneity, with a set of 4 markers used for epithelial and stromal segmentation to support cell-based molecular quantitation.

GE will continue to pursue synergy between two acquired companies – Clarient with its pathology-based technologies and SeqWright with its next generation sequencing and array technologies – to further develop and refine single-marker diagnostics as companions for personalized therapeutics. One specific application they are pursuing is the diagnosis of Alzheimer's disease prior to the development of clinical symptoms (through genetic markers, computer-aided cognitive testing, biospecimen or imaging diagnostics) and the matching of disease-modifying therapeutics specific to the patient status. To take advantage of phenotype data in real-time, GE is also working to develop a common analytics platform that takes advantage of electronic medical records, social networking, and cloud computing to ensure their precision diagnostics are central players as personalized therapeutic products and patient care expand.

Dietrich Stephan (Pitt) began with the observation that "every disease is genetic" and his vision that every patient will serve as their own drug factory (e.g., growing tumors in mice to see what therapeutics are most effective). He then showed as an example of how genomic medicine could be implemented on a population scale screen shots from the company he co-founded, Navigenics. He used these to illustrate the importance of and consumer-friendly approaches to communication (explaining estimated lifetime risk in terms of individual versus population), decision support (distinguishing environmental from genetic contributions to disease and risk), and behavior change (direct-to-consumer genetic testing did not have measurable impact in the short term on psychological health, diet, exercise, or health screening).

He moved on to the need to develop "clinical grade" (CLIA-approved) next-generation sequencing with sensitivity as high as Sanger sequencing. His company Silicon Valley Biosciences partnered with the Mayo Clinic Center for Individualized Medicine to develop automated interpretation engines for clinical genomic interpretation. In synthetic patients, they achieved 100% sensitivity for standard SNVs and indels and 96.4% sensitivity for highly problematic regions. In cancer patients with known mutations, they achieved 100% sensitivity for both standard SNVs and indels, and in fully Sanger sequenced cancer patients, they achieved 99.8% sensitivity. In predicting the pathogenicity of variants of unknown significance, they achieved 93% sensitivity and 97% specificity (versus 79% and 98% for SNPeff and 65% and 94% for SIFT, respectively).

He closed by presenting the future of clinical testing through the example of Mayo Medical Laboratories, which perform over 23 million tests per year for more than 6,600 clients around the world plus education and training and quality assurance programs. He noted the gains to date and gains to be made in terms of laboratory testing market maturity for Mendelian diseases (most mature), anti-cancer therapy,

pharmacogenomics, complex chronic disease, and host-graft/host-pathogen interactions. He predicted that, as with other clusters of biotech industry, a gene science "silicon valley" would soon evolve and suggested that Pittsburgh could have the life sciences and computational resources to support such an industry.

Chris Austin (NCATS) introduced the mission of the National Center for Advancing Translational Sciences and focused his talk on the program to re-engineer translational sciences, with a special focus on the Division of Preclinical Innovation, which seeks both to evaluate existing methods and processes for preclinical research and develop new approaches and technologies in collaboration with NIH intramural and extramural investigators, the FDA, the EPA, and industry. He emphasized the need for improvement in the throughput of the target-to-lead stage of drug development at a logarithmic rather than arithmetic scale.

He started with an overview of the NIH Chemical Genomics Center (NCGC), which had its start in the NIH Common Fund Molecular Libraries program and covers high-throughput screening, chemical informatics, medicinal chemistry, chemical genomics, target-to-lead, toxicology assessment, and siRNA probes. The Center has more than 200 collaborators around the world and focuses on unprecedented targets and rare and neglected diseases. The Center has already made significant contributions, such as quantitative high-throughput screening, which dramatically reduces false negatives (and 4-6 months of development time) by generating a concentration-response curve for each compound over a range of 2 nM to 10μ M, and the discovery of small molecule human PKM2 activators for cancer, an effort that involved the collaboration of multiple labs. The Center also created the NGCG Pharmaceutical Collection (NPC), a definitive database of all known drugs registered or approved for human use and a physical collection of small molecules amenable to high-throughput screening. Currently, the NPC is being screened against Lilly's Phenotypic Drug Discovery Panel to identify novel mechanisms for how drugs work in complex human systems; results will be made freely available online.

Screening for novel drug combinations represents another major initiative. Dr. Austin noted that a successful combinatorics study would require a high-value library of small molecules, and effective plating process, and automated data analysis. Based on results for a single agent, a 6x6 matrix is generated using Mechanism Interrogation Plates (exploits acoustic dispensing technology and advanced informatics) to uncover potential synergies, with good combinations expanded to 10x10 matrix heatmaps to confirm synergistic combinations and perform self-crosses to provide context for activities.

Dr. Austin then moved on to the grand challenge: predicting toxicity. He described a process in which a virtual rat is separated into biochemical, molecular pathway, cellular, tissue, and whole organism components that are tested individually and put back together. The Tox21 Consortium includes NCATS and NCGC, the FDA, the EPA, and NIEHS and the National Toxicology Program—each of which brings

complementary resources and areas of expertise. The goal is to identify patterns of compound-induced biological response and develop predictive models of biological response in humans. Assays are reviewed, optimized, validated, and miniaturized by the Consortium, and results are deposited in the Tox21 chemical inventory, which has over 10,000 plated compound solutions spanning more than 8300 unique chemical substances, and other public databases: EPA ACToR (Aggregated Computational Toxicology Resource), NIEHS CEBS (Chemical Effects in Biological Systems;), and NLM PubChem. At the heart of the Tox21 process is the robotic screening system that makes rigorous high-throughput screening and validation possible (10,000 replicates per experiment), including the screening of the 10,000-compound library 3 times at 15 concentrations per assay. In the toxicant "matrix", cells are treated with a toxicant paired with each of ~70 modulator drugs with known targets; modulators that increase or decrease toxicity give insights into mechanism and inform queries of gene expression.

Currently in Phase II of its research program, Tox 21 is focused on characterizing human variability in *in vitro* responses at both the individual and the population level. For the latter, a collaborative project (TOXICOgenETICS Project) of NCATS, NIEHS, and UNC is testing 8 concentrations of 179 common chemicals (pharmaceutical and environmental) in 1086 cell lines (35% European, 26% Asian, 25% African, 14% Latino) with 1-3 plate replicates. They will turn to crowdsourcing to develop models using the ~2.4 million data points and ~2.5 million SNPs to build models to better predict toxicity.

Phase III will move on to high-content assays and high-throughput gene expression platforms using cells capable of xenobiotic metabolism as well as differentiated embryonic and induced pluripotent stem cells. Goals will include integrating metabolite prediction models as part of the hazard prediction model and expanding testing into model organisms (zebrafish, C. elegans). In parallel with this effort, NCATS has partnered with DARPA on the Microphysiological Systems Program, also known as organ on a chip for drug screening. Twelve awardees are developing 3D microfluidic platforms capable of metabolically recapitulating the cellular complexity of the liver, heart, lung, blood-brain barrier, and other tissues, while another seven awardees are exploring how to differentiate embryonic and induced pluripotent stem cells into multiple cell types capable of representing the cellular architecture within organ systems and thus populating the 3D tissue chips. A long-term goal will be to use these chips as personal drug-testing apparatus. Part of the discussion at this point addressed the inadequacy of testing toxicity predictions generated through these elegant experimental methods, particularly human-based cells and "tissue", in animal models, with an alternative approach applying the iterative QSP process to validate overlapping modules.

In the evening, Ashley Dombkowski (Bay City Capital) led the group through an after-dinner experiment (genetic variant to taste versus not taste paper) to demonstrate the importance of implementing personalized medicine. The difficulty

in getting there comes from shrinking resources as patents expire, as acquisition replaces pipeline development as the source of new molecules, and as venture capital in the life sciences declines. She looked ahead to new trends and paradigms on the horizon, including patients as consumers to be wooed, as by the retail industry and gave examples from Cleveland Clinic (architecture and patient gowns) and Sanofi (The Dx: The Diabetes Experience). She then went through several examples from 23andMe illustrating how a consumer service could be used for research (virtual studies using big data) and patient care. In light of all this, she envisioned the future model for venture capital being focused on those engaged in pipeline development who could demonstrate an obvious breakthrough therapy in a protected therapeutic class using new types of data (addressing a need versus creating demand). If successful, such a start-up would be attractive for purchase by big pharma, fulfilling the dual strategies of improving patient care while supporting a financial exit.

On the second day of the workshop, the Secretary of Health for the Commonwealth of Pennsylvania, Michael Wolf, gave comments on his experience at Pfizer (demise of the blockbuster model), the goals of the state in personalized medicine (leader in technology and innovation, improved home health care), the need for better communication and behavior change, and the role of the state in supporting research and development (review of tobacco settlement funding and new opportunities).

John Wikswo (Vanderbilt) gave a integrated PowerPoint presentation summarizing the discussion at his lunch table and the workshop themes overall. He started with the opening observation raised by their group: We are missing a breadth of clinical data. The diversity of human response is not collected.

Over the next several minutes, he summarized the 3 major categories of bottlenecks in the implementation of QSP: data (do we have the right kind?), technology (do we need more?), and process (legal, data, sociology). With regard to getting enough data of the right kind, we need to select another problem, address the technology issue, or obtain more data using existing approaches (generation, quantification, mining, correlation, interpretation – with a detailed review of the process and challenges for each). With regard to technology, he urged the termination of development of technologies for problems that have already been solved and focus instead on new hardware (wet and dry for biology, chemistry, biochemistry, bioengineering, and computers), software (modeling and computational), and both (adaptive trial design, automated experimental design, automated model inference, closed-loop model-based control of biology). With regard to process, he reviewed legal (consent, IP, liability, right to reconnect), data (standardization, annotation, accessibility, sharing, ownership, curation, incidental findings), and sociology (cooperation, culture, language, silos) bottlenecks that will need to be addressed.

In the end, he concluded that solving the problems of implementing QSP is not as difficult as the problems QSP seeks to solve.

Dr. Shoichet then summarized his group's discussion as focused on actionable points, such as a curated library of well-tested bioactive small molecules with no known toxicities; he assumed about 10-12K such molecules were sequestered away throughout industry but would be incredibly valuable to researchers, as would phenotypic screening libraries. He noted the value of having a brokering process to help align the interests of academia and industry to take advantage of the vertical integration available within industry and cited the Novartis Presidential Postdocs as an example of making the most out of limited resources. He also felt the Pittsburgh Workshop model should be continued so these discussions could be continued, particularly during the lean times, so that action could happen rapidly based on this careful planning when resources became available (gave the example of the Russian film maker Eisenstein). Later in the discussion, Dr. Stevens noted that industry was on board with sharing their compounds but that they needed an incentive to distribute these in an unstructured way to someone who might collect data of questionable quality; he also mentioned that Lily had signed agreements with many universities but that some balked over requests to review manuscripts prior to publication (no ability to block publication – just review prior to submission).

Randy Smith (Pitt) said his group also discussed practical issues related to models of working partnerships among industry, health care, and academia, such as a center for cooperative medicine. He noted that when industry came to meet with UPMC, they were unaware of the possibilities available through such large, technologically advanced health care systems, so the question was how to raise awareness and plan efforts to exploit these resources to advance QSP, big data management, and PM. While the Pittsburgh Workshop was a good start, additional representatives from government, especially the FDA (the FDA representative could not attend due to the government shutdown), would be important.

Jeff Brodsky (Pitt) noted that his group was an academic table who felt sharing among researchers, at least at Pitt, was not a bottleneck, but that collaborating outside the university was due to constraints and delays in the university legal process.. While he had seen academic and industry researchers talk together, he had never seen researchers, fiscal administrators, and attorneys all at the same table to address in a joint discussion the legal restrictions to collaboration. Dr. Shoichet further noted that everything is worked out as one small deal at a time, such that compounds are given out piecemeal, with some never being pursued. Dan Gallahan (NCI) suggested standardized agreement language and tools be developed (versus each institution using its own), and Dr. Smith suggested the state play a role by supporting legislation to provide a platform for sharing. Dave Whitcomb (Pitt) noted that conflict of interest policies governing interactions between physicians and industry at academic medical centers was driving pharma to work with non-academic physicians, thus losing an important link in the research process.

In further discussion about FDA regulatory models, Dr. Berg indicated his support for the FDA position on pharmacogenomics, which allows industry to collect data they might not understand at the time, and Dr. Gallahan confirmed that the FDA was working with NCI to move things forward in this arena. Dr. Brodsky suggested that the FDA might add a portal to help academics learn more about the standardized assays used in testing.

Some discussion focused on scientific process, such as the need for causal modeling (rather than drawing correlations) and dynamical modeling, which is not possible with large datasets of single or limited numbers of measurements. The dynamic modeling in particular would have tremendous cost-savings benefits through the early and rapid identification of which patients will have adverse events and which will not benefit from a particular drug. Dr. Smith noted in that if we could eliminate the ~30% of drugs that don't work (through QSP and PM analyses and applications), we would reduce total healthcare spending by 10%.

The importance of communication both among research silos and with the public was also a common thread. More public relations messages will be needed to explain the added value of personalized medicine, particularly the potential to influence the public with regard to complying with medications and health recommendations and to proactively prevent the sort of social media campaign that damaged immunization rates. Gameification, crowd sourcing challenges (e.g., DREAM, hackathons), and engagement of patients as citizen scientists whose participation can benefit others (23&me examples) were discussed as novel approaches. Participating and giving something back is sufficient motivation for many (i.e., no financial incentive necessary).

Overall, the workshop demonstrated that moving QSP beyond translational research to personalized medicine will require more cultural changes than technological advances. Greater sharing, cooperation, communication, and aligned incentives will have the greatest impact on the successful implementation of QSP in PM.

Initiatives to apply QSP approaches to precision drug discovery and therapy in the PM realm have two complementary needs: QSP scientists rely on clinicians to input the right data (electronic medical record and biospecimens), and clinicians rely on QSP scientists to make sense of the vast amounts of diverse data. Thus, up-front emphasis on the importance of quality control in data input and management is critical to maintain confidence in downstream research and applications.

Pitt and UPMC emerged as a model system in the QSP-PM continuum, as evidenced by their creation of a local ecosystem to support The Cancer Genome Atlas in both research and clinical care that can be expanded to absorb other sources of omic and modeling data generated through QSP approaches. One suggestion was that Pitt-UPMC and Carnegie Mellon University work with industry to identify the problems that Pittsburgh is best suited to tackle (since they cannot do it all, all at once).

Next steps include planning future workshops, perhaps with an alternating focus on QSP and PM, perhaps with more of a focus on science. QSP could become its own Gordon Research Conference as well. The benefit of a Pittsburgh-organized conference is its neutrality and avoidance of expectations (e.g., RFA following an NIH conference, regulations following an FDA conference) and the ability to address culture and process as well as science. The key is that no one agency or organization can solve these problems alone, and that providing a framework for cooperation and advancement of QSP and PM will be essential.

Literature Cited in Presentations and Discussions

Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, Austin CP, Shinn P, Simeonov A, Tice RR, Xia M. The Tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. Drug Discov Today. 2013 Aug;18(15-16):716-23. doi: 10.1016/j.drudis.2013.05.015.

Baker M. Gene data to hit milestone. Nature. 2012 Jul 18;487(7407):282-3. doi: 10.1038/487282a.

Barkan ID. Industry invites regulation: the passage of the Pure Food and Drug Act of 1906. Am J Public Health. 1985 Jan;75(1):18-26.

Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, Ebright RY, Stewart ML, Ito D, Wang S, Bracha AL, Liefeld T, Wawer M, Gilbert JC, Wilson AJ, Stransky N, Kryukov GV, Dancik V, Barretina J, Garraway LA, Hon CS, Munoz B, Bittker JA, Stockwell BR, Khabele D, Stern AM, Clemons PA, Shamji AF, Schreiber SL. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell. 2013 Aug 29;154(5):1151-61. doi: 10.1016/j.cell.2013.08.003.

Berg, J, Rogers ME, and Lyster PM. Systems Biology and Pharmacology. Clinical Pharmacology and Therapeutics. Jul; 88 (1): 17-19.

Berger SI, Ma'ayan A, Iyengar R. Systems pharmacology of arrhythmias. Sci Signal. 2010 Apr 20;3(118):ra30. doi: 10.1126/scisignal.2000723.

Bhattacharyya RP, Reményi A, Yeh BJ, Lim WA. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. Annu Rev Biochem. 2006;75:655-80.

Bilal E, Dutkowski J, Guinney J, Jang IS, Logsdon BA, Pandey G, Sauerwine BA, Shimoni Y, Moen Vollan HK, Mecham BH, Rueda OM, Tost J, Curtis C, Alvarez MJ, Kristensen VN, Aparicio S, Børresen-Dale AL, Caldas C, Califano A, Friend SH, Ideker T, Schadt EE, Stolovitzky GA, Margolin AA. Improving breast cancer survival analysis through competition-based multidimensional modeling. PLoS Comput Biol. 2013 May;9(5):e1003047. doi: 10.1371/journal.pcbi.1003047.

Bloss CS, Schork NJ, Topol EJ. Effect of direct-to-consumer genomewide profiling to assess disease risk. N Engl J Med. 2011 Feb 10;364(6):524-34. doi: 10.1056/NEJMoa1011893.

Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, Huang SS, Schmitz RJ, Urich MA, Kuo D, Nery JR, Qiao H, Yang A, Jamali A, Chen H, Ideker T, Ren B, Bar-Joseph Z, Hughes TR, Ecker JR. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. Elife. 2013 Jun 11;2:e00675. doi: 10.7554/eLife.00675. Print 2013.

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007;2(10):2366-82.

Dutkowski J, Kramer M, Surma MA, Balakrishnan R, Cherry JM, Krogan NJ, Ideker T. A gene ontology inferred from molecular networks. Nat Biotechnol. 2013 Jan;31(1):38-45.

Ernst J, Beg QK, Kay KA, Balázsi G, Oltvai ZN, Bar-Joseph Z. A semi-supervised method for predicting transcription factor-gene interactions in Escherichia coli. PLoS Comput Biol. 2008 Mar 28;4(3):e1000044. doi: 10.1371/journal.pcbi.1000044.

Ernst J, Plasterer HL, Simon I, Bar-Joseph Z. Integrating multiple evidence sources to predict transcription factor binding in the human genome. Genome Res. 2010 Apr;20(4):526-36. doi: 10.1101/gr.096305.109.

Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z. Reconstructing dynamic regulatory maps. Mol Syst Biol. 2007;3:74.

Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, Can A, Corwin A, Dinn S, Filkins RJ, Hollman D, Kamath V, Kaanumalle S, Kenny K, Larsen M, Lazare M, Li Q, Lowes C, McCulloch CC, McDonough E, Montalto MC, Pang Z, Rittscher J, Santamaria-Pang A, Sarachan BD, Seel ML, Seppo A, Shaikh K, Sui Y, Zhang J, Ginty F. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc Natl Acad Sci U S A. 2013 Jul 16;110(29):11982-7. doi: 10.1073/pnas.1300136110.

Grewal A, Stephan DA. Diagnostics for personalized medicine: what will change in the era of large-scale genomics studies? Personalized Medicine. 2013 Nov;10(8):835-848. doi 10.2217/pme.13.82.

Hahn WC, Weinberg RA. Modelling the molecular circuitry of cancer. Nat Rev Cancer. 2002 May;2(5):331-41.

Hidvegi T, Ewing M, Hale P, Dippold C, Beckett C, Kemp C, Maurice N, Mukherjee A, Goldbach C, Watkins S, Michalopoulos G, Perlmutter DH. An autophagy-enhancing drug promotes degradation of mutant alpha1-antitrypsin Z and reduces hepatic fibrosis. Science. 2010 Jul 9;329(5988):229-32. doi: 10.1126/science.1190354.

Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nat Methods. 2013 Nov;10(11):1108-15. doi: 10.1038/nmeth.2651.

Hopkins AL, Mason JS, Overington JP. Can we rationally design promiscuous drugs? Curr Opin Struct Biol. 2006 Feb;16(1):127-36.

Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, Nguyen DT, Austin CP. The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. Sci Transl Med. 2011 Apr 27;3(80):80ps16. doi: 10.1126/scitranslmed.3001862.

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A. 2006 Aug 1;103(31):11473-8.

Iyengar R, Zhao S, Chung SW, Mager DE, Gallo JM. Merging systems biology with pharmacodynamics. Sci Transl Med. 2012 Mar 21;4(126):126ps7. doi: 10.1126/scitranslmed.3003563.

Kanekiyo M, Wei CJ, Yassine HM, McTamney PM, Boyington JC, Whittle JR, Rao SS, Kong WP, Wang L, Nabel GJ. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. Nature. 2013 Jul 4;499(7456):102-6. doi: 10.1038/nature12202.

Kanfi Y, Naiman S, Amir G, Peshti V, Zinman G, Nahum L, Bar-Joseph Z, Cohen HY. The sirtuin SIRT6 regulates lifespan in male mice. Nature. 2012 Feb 22;483(7388):218-21. doi: 10.1038/nature10815.

Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. Nat Biotechnol. 2007 Feb;25(2):197-206.

Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kuijer MB, Matos RC, Tran TB, Whaley R, Glennon RA, Hert J, Thomas KL, Edwards DD, Shoichet BK, Roth BL. Predicting new molecular targets for known drugs. Nature. 2009 Nov 12;462(7270):175-81. doi: 10.1038/nature08506.

Kitano H. A robustness-based approach to systems-oriented drug design. Nat Rev Drug Discov. 2007 Mar;6(3):202-10.

Kokel D, Bryan J, Laggner C, White R, Cheung CY, Mateus R, Healey D, Kim S, Werdich AA, Haggarty SJ, Macrae CA, Shoichet B, Peterson RT. Rapid behavior-based identification of neuroactive small molecules in the zebrafish. Nat Chem Biol. 2010 Mar;6(3):231-237.

Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov. 2004 Aug;3(8):711-5.

Kuo D, Tan K, Zinman G, Ravasi T, Bar-Joseph Z, Ideker T. Evolutionary divergence in the fungal response to fluconazole revealed by soft clustering. Genome Biol. 2010;11(7):R77. doi: 10.1186/gb-2010-11-7-r77.

Laggner C, Kokel D, Setola V, Tolia A, Lin H, Irwin JJ, Keiser MJ, Cheung CY, Minor DL Jr, Roth BL, Peterson RT, Shoichet BK. Chemical informatics and target identification in a zebrafish phenotypic screen. Nat Chem Biol. 2011 Dec 18;8(2):144-6. doi: 10.1038/nchembio.732.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006 Sep 29;313(5795):1929-35.

Learner J. Corporate Venturing. Harvard Business Review. 2013 Oct. http://hbr.org/2013/10/corporate-venturing/ar/1

Lemieux GA, Liu J, Mayer N, Bainton RJ, Ashrafi K, Werb Z. A whole-organism screen identifies new regulators of fat storage. Nat Chem Biol. 2011 Apr;7(4):206-13. doi: 10.1038/nchembio.534.

Lin H, Sassano MF, Roth BL, Shoichet BK. A pharmacological organization of G protein-coupled receptors. Nat Methods. 2013 Feb;10(2):140-6. doi: 10.1038/nmeth.2324.

Ma'ayan A, Jenkins SL, Goldfarb J, Iyengar R. Network analysis of FDA approved drugs and their targets. Mt Sinai J Med. 2007 Apr;74(1):27-32.

Navlakha S, Bar-Joseph Z. Algorithms in nature: the convergence of systems biology and computational thinking. Mol Syst Biol. 2011 Nov 8;7:546. doi: 10.1038/msb.2011.78.

Nussinov R, Tsai CJ. Allostery in disease and in drug discovery. Cell. 2013 Apr 11;153(2):293-305. doi: 10.1016/j.cell.2013.03.034.

Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, Van Deun K, Smith P, Berger B, Heller A. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol. 2000 Aug;32(1):56-67.

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010 Mar;9(3):203-14. doi: 10.1038/nrd3078.

Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. PLoS One. 2007 Mar 21;2(3):e308.

Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordoñez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010 Jan 14;463(7278):184-90. doi: 10.1038/nature08629.

Rogers M, Lyster P, Okita R. NIH support for the emergence of quantitative and systems pharmacology. CPT Pharmacometrics Syst Pharmacol. 2013 Apr 10;2:e37. doi: 10.1038/psp.2013.13.

Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nature Reviews Drug Discovery 3, 353-359 (April 2004) | doi:10.1038/nrd1346.

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T. A travel guide to Cytoscape plugins. Nat Methods. 2012 Nov;9(11):1069-76. doi: 10.1038/nmeth.2212.

Sarewitz D. Science's rightful place is in service of society. Nature. 2013 Oct 31;502(7473):595. doi: 10.1038/502595a.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011 Feb 1;27(3):431-2. doi: 10.1093/bioinformatics/btq675.

Sorger PK, Allerheiligen SRB, et al. Quantitative and Systems Pharmacology in the Post-genomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic Mechanisms. An NIH White Paper by the QSP Workshop Group – October, 2011. http://www.nigms.nih.gov/News/Reports/201110-syspharma.htm

Sutherland JJ, Raymond JW, Stevens JL, Baker TK, Watson DE. Relating molecular properties and in vitro assay results to in vivo drug disposition and toxicity outcomes. J Med Chem. 2012 Jul 26;55(14):6455-66. doi: 10.1021/jm300684u.

Taylor, DL. A new vision of drug discovery and development. European Pharmaceutical Review. 2012 17 (6):

Tice RR, Austin CP, Kavlock RJ, Bucher JR. Improving the Human Hazard Characterization of Chemicals: A Tox21 Update. Environ Health Perspect. 2013. doi:10.1289/ehp.1205784 http://ehp.niehs.nih.gov/1205784/

Tomkins GM. The metabolic code. Science. 1975 Sep 5;189(4205):760-3.

Tsimberidou AM, Iskander NG, Hong DS, Wheler JJ, Falchook GS, Fu S, Piha-Paul S, Naing A, Janku F, Luthra R, Ye Y, Wen S, Berry D, Kurzrock R. Personalized medicine in a phase I clinical trials program: the MD Anderson Cancer Center initiative. Clin Cancer Res. 2012 Nov 15;18(22):6373-83. doi: 10.1158/1078-0432.CCR-12-1627.

Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS, Zhou T, Schmidt SD, Wu L, Xu L, Longo NS, McKee K, O'Dell S, Louder MK, Wycuff DL, Feng Y, Nason M, Doria-Rose N, Connors M, Kwong PD, Roederer M, Wyatt RT, Nabel GJ, Mascola JR. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. Science. 2010 Aug 13;329(5993):856-61. doi: 10.1126/science.1187659.

Yarden Y. The biological framework: translational research from bench to clinic. Oncologist. 2011;16 Suppl 1:23-9. doi: 10.1634/theoncologist.2011-S1-23.

Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. Nat Biotechnol. 2007 Oct;25(10):1119-26.

Zhao S, Nishimura T, Chen Y, Azeloglu EU, Gottesman O, Giannarelli C, Zafar MU, Benard L, Badimon JJ, Hajjar RJ, Goldfarb J, Iyengar R. Systems pharmacology of adverse event mitigation by drug combinations. Sci Transl Med. 2013 Oct 9;5(206):206ra140. doi: 10.1126/scitranslmed.3006548.

Zinman G, Brower-Sinning R, Emeche CH, Ernst J, Huang GT, Mahony S, Myers AJ, O'Dee DM, Flynn JL, Nau GJ, Ross TM, Salter RD, Benos PV, Bar Joseph Z, Morel PA. Large scale comparison of innate responses to viral and bacterial pathogens in mouse and macaque. PLoS One. 2011;6(7):e22401. doi: 10.1371/journal.pone.0022401.

Zinman GE, Naiman S, Kanfi Y, Cohen H, Bar-Joseph Z. ExpressionBlast: mining large, unstructured expression databases. Nat Methods. 2013 Oct;10(10):925-6. doi: 10.1038/nmeth.2630.