ABSTRACT — Biomarkers are invaluable drug development tools to assess and monitor safety in early clinical trials especially when exposure margins are limiting for promising therapeutics. Although progress has been made towards identifying and implementing translational safety biomarkers for a number of organ toxicities such as kidney and liver, significant biomarker gaps still exist to monitor toxicities for testis, pancreas, etc. Several precompetitive consortia [e.g., Predictive Safety Testing Consortia (PSTC), Innovative Medicines Initiative (IMI)] are working with industry, academia, government, patient advocacy groups and foundations with a goal to qualify biomarkers such that they can be used in preclinical studies and clinical trials to accelerate drug development. This manuscript discusses the complexities of novel biomarker discovery, validation and international regulatory qualifications intended for clinical trial applications and shares specific examples from Pfizer Research and Development. As safety biomarkers become widely accepted and qualified by the regulatory agencies, they will increasingly be implemented in early clinical trials, play a key role in decision making and facilitate the progression of promising therapeutics from preclinical through clinical development.

Key words: Safety biomarker, Clinical translation, Biomarker qualification

INTRODUCTION

The pharmaceutical industry continues to face challenges in terms of declining productivity. The overall cost of drug development and time for most drugs to reach market has been increasing over the past several decades. From 1975 to 2015, the total expenditures per year for drug discovery and development rose approximately 5-fold while the number of United States Food and Drug Administration (FDA) new drug registrations per year remained relatively flat (Boyer et al., 2016). Substantial resources are being invested in research and development (R&D) across the industry into compounds that eventually fail. The current pharmaceutical R&D process results in only ~10% of the molecules entering phase 1 clinical trials reaching full approval by the FDA (Hay et al., 2014). This high rate of attrition is a key driver in reducing productivity and a significant challenge to the industry. There are a number of explanations for the decrease in productivity including a focus of the industry in areas of unmet medical need and novel biological mechanisms with high risk of failure, a higher entry bar for new drugs due to a competition with enhanced standard of care, higher regulatory hurdles, commercial and financial portfolio decisions and an increase in the complexity and cost of clinical trials (Roberts et al., 2014; Hay et al., 2014); however when a root-cause analysis was completed on 359 phase 3 and 95 new molecular entities and biologic license applications, safety and efficacy were shown to be the two primary causes for compound suspension (Hay et al., 2014).

The standard preclinical testing paradigm has markedly improved drug safety over the past 30 years. In 2000, the results from a multinational pharmaceutical survey and the outcome of an International Life Sciences Institute (ILSI) Workshop were reported and showed that 70% of human toxicity observed during clinical trials is predicted by preclinical studies (Olson et al., 2000). In addition, a 2013 study in Japan reported that 48% of adverse drug reactions observed in clinical trials were predicted by a comprehensive preclinical safety assessment (Ahuja and Sharma, 2014); and in 2017, the IQ Consortium created and utilized an industry-wide nonclinical to clinical translational database to report that animal studies not only have value in predicting human toxicities but also that an absence of toxicity in nonclinical studies predicts a similar outcome in the clinic (Monticello et al., 2017). However, the inability to predict failures before or early in clinical trials remains a main cause of
attrition (Ahuja and Sharma, 2014) as failures in clinical safety not predicted in preclinical studies continue to be a major cause of compound termination (Clark and Steger-Hartmann, 2018). In addition, clinical failures due to toxicity are not limited to the early clinical portfolio but are also evident in late stage development and post-marketing surveillance (Waring et al., 2015). Thus while preclinical data has had a significant impact on improving clinical drug safety, failures due to toxicity remain a key challenge facing the industry.

The high rate of attrition and cost of drug development has prompted a surge in research in the area of biomarkers. According to the FDA/National Institute of Health (NIH)’s BEST (Biomarkers, Endpoints, and other Tools Resource) guide, a biomarker is a ‘defined characteristic that is measured as an indicator of normal biologic processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions (FDA, 2016a).’ Biomarkers are invaluable drug development tools utilized to better understand the translatability of preclinical findings into human adverse events and to better assess and monitor these findings in the clinic especially when exposure margins are limiting for promising therapeutics. Although progress has been made towards identifying and implementing translational safety biomarkers for a number of organ toxicities such as kidney and liver, significant biomarker gaps still exist. Several precompetitive consortia (e.g., PSTC, IMI) are working with industry, academia, government, patient advocacy groups and foundations with a goal to qualify biomarkers such that they can be implemented in preclinical studies and clinical trials to accelerate drug development. This paper discusses the complexities of discovering, validating and qualifying novel biomarkers to understand organ toxicities and to accelerate drug development decisions. Specific biomarker case examples are highlighted to describe their implementation in preclinical and early clinical development to characterize safety issues, understand mechanisms and monitor clinical safety. As safety biomarker assays are validated and qualified by the regulatory agencies, they will increasingly play a key role in decision making and facilitate the progression of promising therapeutics from preclinical through clinical development. This, however, is not a trivial endeavor and the path to success may best be achieved through a collaborative effort among industry, academia and regulatory partners.

BIOMARKERS

In 2004, the FDA’s Critical Path Initiative emphasized the need for innovation in drug development and suggested the use of biomarkers to evaluate and predict safety and effectiveness, provide informative links between mechanism of action and clinical utility, understand the translatability between preclinical species and humans, and serve as surrogate endpoints. With a focus on generating and implementing translatable biomarkers early in the drug development process, the goal was for the pharmaceutical industry to improve the high attrition rates and process inefficiencies in terms of cost and time observed across the industry (Haim, 2011; Woodcock and Woosley, 2008). In drug development, biomarkers categorized as target biomarkers assess the ability of the drug to reach and engage the target while mechanism biomarkers test the ability of the drug to induce the expected measurable molecular or cellular pharmacodynamics response. Finally, the linkage of the mechanism to the clinical response or efficacy is assessed by disease biomarkers or clinical activity scores. Ideally, safety biomarkers inform the presence or extent of toxicity and thus increase confidence in safety and the ability to predict, detect and monitor the progression of drug-induced toxicity. The ability of safety biomarkers to detect early toxicity, monitor onset and reversibility, and manage adverse effects observed in the clinic will determine its overall utility and impact in preclinical and clinical drug development. With the emergence of innovative technologies and a diverse set of in vitro and in vivo models, the development of novel safety biomarkers has evolved to include not only the singular measurement of circulating proteins but also expression profile signatures in blood/tissue, noninvasive imaging and genetic variations in DNA (Currid and Gallagher, 2008). Assays for these biomarkers range from exploratory fit for purpose to fully validated and qualified by the regulators for a specific context of use and the process of assay validation should be regarded as being continuous and evolving over-time. Pfizer has a strong commitment to developing and implementing translatable biomarker strategies within drug development including applying protein biomarkers for drug-induced toxicity in the stomach, kidney, skeletal muscle and liver and small molecules/metabolites as biomarkers for drug-induced toxicity to the liver and heart. Several biomarker case examples from Pfizer R&D efforts are included in this review. All procedures performed on animals used in Pfizer examples were in accordance with regulations and established guidelines and were reviewed and approved by an Institutional Animal Care and Use Committee.
APPLICATION OF SAFETY BIOMARKERS

Example 1: Translational kidney biomarker panel as a preclinical screening tool

A number of emerging urinary kidney markers including kidney injury molecule 1 (KIM-1), clusterin, microalbumin, trefoil factor 3, α-glutathione S-transferase, N-acetyl-β-D-glucosaminidase (NAG), neutrophil gelatinase-associated protein (NGAL) and osteopontin are currently being evaluated across the industry. Published studies of emerging markers show promising results in terms of increased sensitivity and specificity in comparison to the standard serum markers, serum creatinine (sCr) and blood urea nitrogen (BUN) (Chen et al., 2017; Dieterle et al., 2010a; Fuchs and Hewitt, 2011; Ozer et al., 2010; Koyner et al., 2010). Although sCr and BUN continue to be utilized as primary renal markers in clinical practice, they are considered poor indicators of early renal dysfunction due to limited sensitivity, remaining relatively unchanged until significant renal damage occurs. Consequently, further evaluation of novel urinary biomarkers of acute kidney injury (AKI) is warranted both in preclinical species and in humans. In order to better understand the performance of KIM-1, NGAL, and NAG in the rat and to provide foundational information for potential clinical studies, a rat study was conducted to compare the performance of these urinary biomarkers to the standard serum markers of AKI for the detection of nephrotoxicity (Burt et al., 2014). These 3 markers were selected after a prescreening of a larger panel based on performance after administration of polymyxin B, a polypeptide antibiotic used for the treatment of life-threatening Gram-negative bacterial infections. Use of these antibiotics in the clinic has been drastically limited due to drug-induced AKI and thus the further development of more-sensitive renal markers has the potential to not only impact drug development but also patient care in the clinic. In this study, rats (5/group) were dosed with 0.1, 0.4, 1, 4 or 10 mg/kg polymyxin B for 2 or 14 days and then necropsied. The highest tolerated dose was 4 mg/kg. Histopathological examination of kidneys from the treated rats showed minimal chronic progressive nephropathy (CPN) in a single rat at the 3 lowest doses, minimal to moderate tubular degeneration/regeneration in 4 animals dosed at 4 mg/kg/day and minimal tubular necrosis in a single rat at 10 mg/kg/day, indicative of drug-induced kidney injury. Of the biomarkers tested, only NGAL and KIM-1 produced dose-dependent statistically significant elevations, 13-fold and 3-fold respectively, as early as 48 hr post-dose. In comparison, sCr was not affected by polymyxin treatment and increases in BUN were small (1.3-fold) and were not dose-dependent. Urinary NGAL, however, was the most sensitive biomarker of AKI in this rat model, responding to the early onset of kidney injury observed with polymyxin. Since urinary NGAL levels reached a maximum increase at 48 hr post-dose (Fig. 1) that correlated with kidney histopathology, a 2-day study design was implemented for screening and rank ordering potential drug candidates utilizing NGAL as the biomarker for AKI. This design provided an important tool for selecting novel polymyxin B analogs with an improved kidney safety profile, affording an efficient, cost-effective drug development biomarker-driven strategy. This work also provided a basis for potential clinical studies to further evaluate the diagnostic utility of these urinary kidney biomarkers in patients (Burt et al., 2014).

Example 2: Stress-associated hormones as mechanism-based biomarkers utilized to mitigate safety risk

Spontaneous histologic changes in laboratory animal species including rabbits can hamper accurate toxicologic interpretation in preclinical safety studies, especially if detailed study procedures are included. As early as 1924, publications described inflammatory heart findings with multifocal myocardial infiltrates of lymphocytes and/or macrophages in otherwise healthy rabbits (Sellers et al., 2017). In order to better characterize these myocardial findings and further understand the impact of study-related procedures on these findings, a large study...
was designed in New Zealand White female rabbits with either an increase (Group 1) or decrease (Group 2) in the number of study-related procedures and animal handling. Blood was collected for coagulation, hematology and clinical chemistry analysis including stress-associated serum biomarkers (epinephrine, norepinephrine, cortisol, and corticosterone) at various time points throughout the study for Group 1 and only at baseline and necropsy for Group 2 to minimize handling and procedures (Sellers et al., 2017). Group 1 animals with the increased procedures had a higher incidence of inflammation with degeneration/necrosis in cardiac myocytes than the animals from the minimal procedure group (Group 2). Serum stress markers including cortisol and norepinephrine showed the greatest response in Group 1 animals with peak values usually occurring 4 hr and 48 hr post-dose respectively (Fig. 2). This study provided further evidence that increased procedures and handling during study conduct exacerbates the frequency and severity of myocardial inflammatory findings and may be mediated by a stress response. Based on these findings, it was proposed that stress hormones, in particular norepinephrine and cortisol, could be utilized to assess the risk of the myocardial findings observed in rabbit toxicology studies (Sellers et al., 2017).

Example 3: Hyaluronic acid as a translational mechanism-based marker of liver microvascular injury

Hyaluronic Acid (HA) is a polysaccharide located in the extracellular matrix. It is synthesized by mesenchymal cells and cleared almost exclusively by liver sinusoidal endothelial cells (SECs). Circulating HA levels have been shown to be elevated with structural and/or functional damage to the liver SEC and have shown promise in the clinic when studied for the noninvasive detection of sinusoidal obstruction syndrome (SOS). Currently, the diagnosis of SOS in the clinic relies on nonspecific clinical and laboratory measures and/or events occurring late in the development of the disease including jaundice, painful hepatomegaly, weight gain and ascites. SOS has been reported in patients treated with antibody-calicheamicin conjugates including gemtuzumab ozogamicin and inotuzumab developed for acute myeloid leukemia and acute lymphoblastic leukemia, respectively (Guffroy et al., 2017). Liver toxicity based on elevated aspartate aminotransferases (AST) and bilirubin levels was observed with occasional hepatic SOS following treatment with these antibody-drug conjugates (ADCs). While these two ADCs have different monoclonal antibodies, they are composed of the same linker and calicheamicin payload. To further evaluate the mechanism of the adverse event and to identify potential safety biomarker to detect the damage throughout the disease course, an experiment was initiated to characterize the liver injury observed with the ADCs. Cynomologous monkeys were dosed with an antibody-calicheamicin conjugate containing the same linker-payload as gemtuzumab ozogamicin and inotuzumab. Monkeys were dosed with up to 3 intravenous bolus injections 3 weeks apart and were necropsied 48 hr following the first dose on day 3 and 3 weeks after the third administration on day 63. Liver histopathology showed midzonal degeneration and loss of SECs on day 3 and variable endothelial recovery and progression to a combination of sinusoidal capillarization and sinusoidal dilation/hepatocellular atrophy, consistent with early SOS. Minimal increases in AST levels (up to 3.1x) were observed on day 4 and remained elevated over the duration of the study. HA levels (up to 8.5x) were elevated on day 3 and were sustained thru day 63 in all treated monkeys (Fig. 3) demonstrating the ability of this marker to detect early structural damage (day 3) and later functional impairment (day 63) of SECs. HA also showed good correlation to AST levels and microscopic liver findings throughout the study. Since HA, unlike AST, is linked mechanistically to SOS, it was proposed as a sensitive exploratory diagnostic marker of liver micro-

**Fig. 2.** Norepinephrine and Cortisol levels in rabbits following dosing with saline on Days 29, 43 and 57 with blood collections pre-dose (vertical lines) and at 4, 24 and 48 hr post-dose. Data represents group means ± SEM.
vascular injury capable of non-invasive detection of SOS in the clinic (Guffroy et al., 2017).

Example 4: Gastrin as a translational safety biomarker for drug-induced stomach toxicity

Gastrin is secreted from the stomach and plays a key role in the regulation of gastric acid secretion. Gastrin is synthesized in special endocrine cells (G cells) primarily in the antral region of the gastric mucosa and binds receptors found predominantly on parietal cells stimulating gastric juice secretion, the best-known of which is hydrochloric acid (HCl) (Henderson, 2001). The capacity of the stomach to secrete HCl normally is directly proportional to the parietal cell number. High levels of circulating gastrin can occur when the pH of the stomach is high while gastrin secretion by antral G cells is inhibited by the direct action of acid on the G cells. When the stomach lining is damaged and unable to produce and release acid, gastrin continues to be secreted from the fundus and circulating levels rise. Changes in serum gastrin have been shown to be predictive of the functional status of the antral mucosa, making it attractive as a potential biomarker of drug-induced stomach toxicity (Graham et al., 2006; Nicolaou et al., 2014). To further evaluate gastrin as a potential marker of stomach toxicity, a 28-day dog study was conducted with a compound associated with gastric effects. Minimal to moderate atrophy and minimal degeneration of the fundic mucosa were seen at histologic examinations. Mean serum gastrin levels in the treated animals were 23x higher than concurrent controls (Fig. 4) and were attributed to the lack of acid production by the damaged parietal cells, indicating a failure of the feedback mechanism that controls the acid output in the stomach. In individual dogs, serum gastrin levels were elevated (up to 52x) compared to concurrent controls and these levels correlated with the severity of the adverse microscopic stomach findings. Based on these data, serum gastrin was proposed as a non-invasive stomach-specific biomarker to monitor for stomach toxicity in the clinic.

PATHWAYS TO BIOMARKER ACCEPTANCE

Many of the safety biomarkers currently considered conventional and measured routinely in both preclinical and clinical drug development as well as in clinical practice were established prior to the implementation of a formal regulatory qualification process. They were accepted based on scientific community consensus, i.e. review of datasets published in peer reviewed journals, experience in clinical practice, and recommendations from professional medical associations. For example, cardiac troponin (cTn) was recognized as a clinical biomarker of acute myocardial infarction by the American College of Cardiology and the European Society of Cardiology in 2000 (Jaffe, 2001) just thirteen years after the development of the first troponin assay (Danese and Montagnana, 2016). Diagnostic criteria were recommended based on pub-
lished literature and clinical experience with an acknowledgment that the criteria should continue to evolve as additional knowledge was gained (Jaffe, 2001). Subsequent to clinical acceptance, cTn was evaluated as a preclinical marker by the Health and Environmental Sciences Institute (HESI) Cardiac Troponins Biomarker Working Group (Reagan, 2010) and received full qualification from the FDA in 2012 based solely on evidence from peer reviewed scientific literature (FDA, 2012).

Another pathway to biomarker approval is through biomarker evaluation for a drug-specific application. In this case, the biomarker may only be used in a single drug development program and the data required to support the use of the biomarker in the program is determined by the sponsor communicating directly with the regulatory reviewing division accountable for the program (Mattes and Goodsaid, 2018). Utilization of this pathway limits the knowledge of the biomarker’s performance and intended use for the specified program but may enable time saving and efficient progression of a promising compound into clinical development.

In order for a biomarker to be approved for multiple drug development programs, it must pass the rigor of the full regulatory qualification process. Qualification is defined by the FDA as ‘a conclusion that within the stated context of use, the results of assessment with a drug development tool (biomarker) can be relied upon to have a specific interpretation and application in drug development and regulatory review (FDA, 2014).’ Biomarkers reaching this milestone can support regulatory decisions in drug development programs for the approved context of use (COU) in clinical trials. According to the FDA, the COU is ‘a comprehensive and clear statement that describes the manner of use, interpretation, and purpose of use of the biomarker in drug development (FDA, 2016b).’ With the enactment of the 21st Century Cures Act in December 2016, an updated multi-stage biomarker qualification process was established which included three submission stages: the Letter of Intent, the Qualification Plan and the Full Qualification Package. The Biomarker Qualification Program is one of the Drug Development Tools created by the Center of Drug Evaluation and Research to provide a framework for development and regulatory acceptance of biomarkers for use in drug development programs. Similar programs outlining processes for the submission and review of data supporting the approval of new biomarkers are also in place at the European Medicines Agency (EMA) and the Japanese Pharmaceutical and Medical Devices Agency (PDMA). While the agencies work closely on qualification efforts, a fully harmonized approach has yet to be established. However, these regulatory pathways provide a process to review, evaluate and adopt new tools into regulatory decision making in drug development and facilitate consensus science and acceptance of the biomarker’s proposed COU in drug development (Dennis et al., 2013).

To better understand the elements of biomarker qualification, a framework for evidentiary standards for biomarker qualification was recently proposed under the auspices of the Foundation for the National Institute of Health (FNIH) Biomarkers Consortium with representatives from FDA, NIH, industry, patient groups and academia. Five components were incorporated into the framework and included: 1) defining a statement of need (knowledge gap or drug development need) that the biomarker intends to address; 2) defining the COU; 3) assessing the benefits in light of the COU; 4) assessing the risks with regards to the COU and 5) defining the evidentiary criteria required to support the COU (Leptak et al., 2017). The COU determines the level of evidence needed both for the validation of analytical technology employed and for the qualification of the biomarker. The greater the risk to human health of an incorrect decision based on the use of the biomarker, the greater the level of evidence required for the qualifying the biomarker for that COU (Dennis et al., 2013). The assessment of the evidentiary criteria is intended to be utilized as a ‘communication tool for gaining alignment between submitters and FDA reviewers at several key milestones for a biomarker development plan: (i) initial discussions to align expectations; (ii) purposeful interim progress updates to ensure that evidence expectations have been met before proceeding further; and (iii) review evaluation to support the qualification outcome (Leptak et al., 2017).’ The ultimate goal of this effort was to improve the quality of submissions to the FDA, facilitate a level of predictability in the qualification process, and provide clarity as to the type and level of evidence needed to support a biomarker’s COU (Leptak et al., 2017).

Due to the complexity of the qualification process and considerable resources required to reach full qualification of a biomarker, efforts are primarily focused in consortia such the HESI, the PSTC, the IMI Safer and Faster Evidence-based Translation (SAFE-T) consortium and more recently the IMI Translational Safety Biomarker Pipeline (TransBioLine) consortium. To date, 4 safety biomarker submissions have been successful. This included 2 preclinical qualifications for a number of emerging urinary nephrotoxicity biomarkers and one nonclinical qualification of circulating cardiac troponins as indicators of cardiotoxicity. Earlier this year, the first clinical safety biomarker submission reached approval for a panel of urinary

S. Schomaker et al.
biomarkers to aid in the detection of kidney tubular injury in phase 1 trials in healthy volunteers (FDA, 2018b).

**BIOMARKER QUALIFICATION EFFORTS**

**Serum glutamate dehydrogenase (GLDH) as a specific biomarker for hepatocellular injury**

GLDH has been shown to be a sensitive measure of hepatotoxicity in preclinical species and in humans (Giffen et al., 2003; Schomaker et al., 2013) and is currently going through the regulatory qualification process sponsored by the PSTC and the Duchenne Regulatory Science Consortia (D-RSC). The gold standard biomarker for the diagnosis of liver injury is alanine aminotransferase (ALT). However since ALT is also present in myocytes, serum ALT activities can increase with muscle injury; thus, the development of more specific biomarkers for drug-induced liver injury (DILI) is needed. For this qualification, the proposed COU for GLDH is that ‘elevated serum GLDH activity is a measure of hepatocellular injury, and can be used in healthy subjects and patients as an adjunct to ALT, the current standard biomarker used to assess hepatocellular injury, in all stages of drug development. In a clinical situation when ALT increases are observed, GLDH can lend weight of evidence to confirm or rule out hepatocellular injury (EMA, 2017).’ The qualification submission will include a full technical validation of the assay and an evaluation of the clinical relevance of the biomarker, e.g., added value relative to aminotransferase activity, correlation to histopathology in a preclinical species and exploratory and confirmatory analyses in clinical subjects with regulatory guidance. The evaluation of clinical relevance includes establishing reference ranges for healthy subjects and evaluating the influence of gender and age, confirming GLDH as a sensitive biomarker of liver injury and establishing medically relevant cutoffs for DILI, and establishing GLDH as a specific biomarker of hepatocellular injury in comparison to ALT. In March of 2017, a joint FDA and EMA biomarker qualification consultation meeting was held for GLDH. This meeting provided regulatory support for using organ injury induced by diseases with a wide range of etiologies as approximation of chemical-induced organ injury for evaluation of performance of novel biomarkers. This was a paradigm shift in the development of safety biomarkers, since it eliminates the need for lengthy clinical trials and improves feasibility and efficiency of biomarker research. The EMA issued a Letter of Support (LoS) in November 2017 demonstrating the Agency’s support of the qualification and provided additional feedback regarding data needed for achieving full qualification of GLDH as a “Drug Development Tool”. This LoS not only recognized the potential value of GLDH as a liver specific biomarker of hepatocellular injury to address important unmet medical need but also endorsed data interpretation including medically relevant levels of GLDH (2.5x and 5x above upper limit of normal) that were established as part of the qualification effort (EMA, 2017). This qualification effort, initiated in 2015 supported by Pfizer internal exploratory data, is expected to reach fruition in 2019 with the submission of the full qualification package. The formal qualification of GLDH as a liver specific biomarker of hepatocellular injury will not only allow for the broad application of GLDH across programs, but will also enable the diagnosis of the onset of liver disease in subjects with underlying muscle impairments, which is an important unmet medical need widely recognized by the medical community. To this end, Pfizer has partnered with the PSTC, the Duchenne Regulatory Science Consortium (D-RSC), Roche, the maker of the research grade GLDH assay, and FDA to validate the GLDH assay as an *in vitro* diagnostic (IVD) which would allow the assay to be utilized in clinical practice and lead to an improved standard of care for patients living with muscle disease.

**Next generation biomarkers for skeletal muscle degeneration**

Serum AST and creatine kinase (CK) have been used for decades as the primary biomarkers for skeletal muscle (SKM) injury as a measure of myocyte degeneration/necrosis. However, these markers lack tissue specificity for SKM and sensitivity for SKM degeneration/necrosis in both rats and humans. In 2010, the PSTC’s Skeletal Muscle Working Group was formed with a goal to identify and qualify novel safety biomarkers of drug-induced SKM injury that would add value to the current markers, CK and AST, for monitoring SKM injury. The muscle injury panel (MIP) selected for evaluation included skeletal troponin I, myosin light chain, fatty acid-binding protein and creatine kinase measured by a mass assay. These markers were assessed in 34 rat studies and were shown to outperform AST and CK (enzymatic assay) individually and as a panel in terms of sensitivity and specificity and/or added value for the diagnosis of drug-induced SKM injury defined as myocyte degeneration/necrosis (Burch et al., 2016). Based on this data, the FDA and EMA issued Letters of Support (EMA, 2015; FDA, 2015) for the use of these markers in preclinical development and encouraged their use in early clinical trials in an exploratory context. This endorsement encouraged Burch et al. (Burch et al., 2015) to evaluate the translatability of these markers in patients with DMD and other
muscular diseases and their ability to monitor disease progression and the response to treatment. In this study, the MIP biomarker responses was compared to current clinical assessments including CK activity, ambulatory status and cardiac function and were shown to not only better reflect the patients’ disease state compared to CK activity but also correlate with clinical endpoints in patients with DMD and other muscular diseases. These preclinical and clinical evaluations provided support for a qualification submission which is currently in progress. This qualification effort, sponsored by the PSTC, will evaluate the marker’s ability to monitor SKM degeneration/necrosis in conjunction with AST and CK enzymatic activity in early clinical trials and encourage the utilization of the MIP markers throughout drug development to improve patient safety in clinical trials (Burch et al., 2016). In support of the qualification effort, Pfizer has included the MIP markers on a number of clinical trials as exploratory biomarkers; however, the utility of the markers as SKM safety or efficacy endpoints to these therapeutic interventions has yet to be determined (Goldstein, 2017).

Kidney safety urine biomarker qualifications

In 2008 the first formal qualification of preclinical safety biomarkers was granted by the FDA (FDA, 2008) and EMA (EMA, 2008a) followed by PMDA in 2010 (PMDA, 2010) for seven urinary safety biomarkers submitted by the PSTC. In 2010, a qualification was rendered by the FDA for the two preclinical urinary biomarkers, clusterin, renal papillary antigen-1 (RPA-1), submitted by HESI and in 2018 the FDA qualified the first clinical safety biomarkers, a set of six urinary markers interpreted as a Composite Measure (CM) (FDA, 2018a) submitted by PSTC. In all cases, the biomarkers are to be used in conjunction with the traditional measures, serum creatinine (sCr) and blood nitrogen urea (BUN), for the evaluation of nephrotoxicity. sCr and BUN are both insensitive and nonspecific, changing only after significant injury and with a time delay relative to the onset of injury (Vaidya et al., 2008) limiting their ability to accurately estimate injury onset and the severity of the dysfunction following injury (Ferguson et al., 2008).

The seven biomarkers included in the PSTC preclinical submission included KIM-1, clusterin (CLU), albumin, total protein, β2-microglobulin, cystatin C and trefoil factor 3 (TFF3) in urine. The submission contained data and data interpretation from a number of rat studies, a review of the scientific literature in humans (KIM-1, albumin, total protein, cystatin C, and β2-microglobulin), COUs for each biomarker (Dieterle et al., 2010b) and key conclusions. The PSTC put forth three specific biomarker claims in the submission. First, urinary KIM-1, CLU, and albumin can individually outperform and add information to BUN and sCr assays as early diagnostic biomarkers of drug-induced kidney tubular alterations in rat toxicology studies. Second, urinary TFF3 can add information to BUN and sCr assays in rat toxicology studies as an early diagnostic biomarker of drug-induced acute kidney injury tubular alterations. Third, total urinary protein, cystatin C, and β2-microglobulin can individually outperform sCr assays and add information to BUN and sCr assays as early diagnostic biomarkers in rat toxicology studies of acute drug-induced glomerular alterations or damage resulting in impairment of kidney tubular reabsorption (Dieterle et al., 2010b). In addition, the PSTC claimed that this rat data taken together with the published peer-reviewed clinical data supported the voluntary use of KIM-1, albumin, total protein, cystatin C, and β2-microglobulin as bridging markers for early clinical trials on a case-by-case basis when concerns are generated in GLP animal toxicology studies. During the qualification effort, the FDA and EMA provided feedback regarding submission gaps, statistical considerations and preliminary conclusion statements (EMA, 2008b) which culminated in letters of acceptance for the preclinical qualifications from both agencies. In 2008, the FDA and EMA concluded that the urinary kidney biomarkers KIM-1, CLU, albumin, total protein, β2-microglobulin, cystatin C and TFF3 were acceptable for the detection of acute drug-induced kidney injury in rats to be included along with traditional clinical chemistry markers and histopathology in toxicology studies. The EMA also stated that while it was worthwhile exploring their utility in early clinical trials as clinical biomarkers, until additional data was available “to correlate the biomarkers with the evolution of the nephrotoxic alternations, and their reversibility, their general use for monitoring nephrotoxicity in clinical setting cannot be recommended (EMA, 2008b).” Based on this data, in 2010 the PDMA announced the first biomarker qualification decision under the new consultation process on pharmacogenomics/biomarkers use in Japan. These recommendations, consistent with the concept of a progressive biomarker qualification, prompted the generation and submission of additional preclinical and clinical data to expand the COU.

In an expanded effort, the SAFE-T and PSTC consortia continued the work on urinary kidney injury biomarkers by evaluating the performance of KIM-1, CLU, NGAL, albumin, total protein, cystatin C, α-glutathione S-transferase and urinary osteopontin in clinical trials. The trials included a study in healthy volunteers, an exploratory study with cisplatin-treated cancer patients, and a study
in patients undergoing coronary angiography. The markers selected are localized in different regions of the nephron thus the panel was expected to respond to a variety of nephrotoxicants. Based on the readout from these studies, both the FDA and EMA issued Letters of Support in 2016 encouraging the exploratory use of these markers as biomarkers of renal tubular injury in early clinical trials to be used in conjunction with traditional biomarkers and clinical and nonclinical findings (FDA, 2016c; EMA, 2016).

Also in 2016, the FNIH Biomarker Consortium and the PSTC submitted a qualification submission for a kidney injury biomarker panel to be interpreted as a Composite Measure (the geometric mean of the fold change from baseline of the six urine biomarkers normalized to urine creatinine) of the following six biomarkers: KIM-1, CLU, cystatin C, NAG, NGAL and urinary osteopontin. According to the COU, “the safety composite biomarker panel is to be used in conjunction with traditional measures to aid in the detection of kidney tubular injury in phase 1 trials in healthy volunteers when there is an a priori concern that a drug may cause renal tubular injury in humans (FDA, 2018a).” The qualification strategy included a nonclinical phase, a clinical exploratory phase and a clinical confirmatory phase with a cisplatin study in cancer patients and an aminoglycoside study in cystic fibrosis patients. Based on this data, this panel, interpreted as a CM, was the first clinical safety biomarker to be qualified by the FDA. Following on the success of this qualification, the FNIIH and PSTC have continued their partnership and have submitted a letter of intent to FDA, EMA and PDMA to further qualify the panel of biomarkers based on the individual biomarker response thus expanding on the qualified CM COU.

**CURRENT STATE OF TRANSLATIONAL SAFETY BIOMARKER DEVELOPMENT**

Significant progress has been made towards implementing translational safety biomarker strategies into the drug development portfolio. Several pharmaceutical companies have been adopting exploratory biomarkers in their preclinical and clinical drug development programs and contributing to consortia led qualification efforts. While much progress has been made towards the qualification of biomarkers for some organ injuries including liver, kidney, and muscle, significant gaps remain for other target organs such as biomarkers of vascular injury in humans, sensitive biomarkers of pancreatic injury, circulating biomarkers of testicular toxicity, and biomarkers of injury to the central nervous system. Because the process of biomarker development is complex and requires considerable resources (Gerlach *et al.*, 2018), these efforts will best be pursued by consortia such as PSTC and IMI. Prior to widespread acceptance, each biomarker must undergo a rigorous assay validation; demonstrate relevance to humans and an association with clinical endpoints reproducibly in multiple studies and gain consensus regarding the level of evidence needed to support a qualification for the stated COU. Due to the complexity and feasibility challenges of prospective randomized drug intervention clinical trials, alternative sample collection strategies like the prospective collection of samples of organ damage etiologies from patients undergoing hospital visits employed in the GLDH qualification effort, need to be considered. These disease patient populations, once identified, can serve as surrogates for drug-induced organ injury to evaluate biomarker performance as long as common molecular and mechanistic pathways are shared between the disease and organ toxicity of interest. The prospective collection and storage of samples from clinical trials conducted for drug development is also an option (Aubrecht *et al.*, 2013). As safety biomarkers become accepted and qualified by the regulatory agencies, they will increasingly play a key role in decision making and facilitate the progression of promising therapeutics from preclinical through clinical development, leading to the realization of our most critical goal: delivering the right dose of the right medicines to the right patients at the right time (Gerlach *et al.*, 2018).

**SUMMARY AND CONCLUSIONS**

Safety and tolerability of newly approved drug candidates still remains a key concern in drug development leading to black box warnings and withdrawals of promising therapeutics. A major hurdle is the lack of validated and qualified safety biomarkers that accurately diagnose, predict and inform mechanism of organ toxicities. Regulatory qualification of a safety biomarker is critical for routine application in clinical development. To address this hurdle and the complexities in biomarker development, considerable resources and partnerships across industry, academia, and regulators is needed. Although tremendous progress has been made towards identifying and implementing translational safety biomarkers for organ toxicities such as kidney and liver, significant gaps still exist to monitor toxicities for other common target organs such as pancreas, central nervous system, testis, and skeletal muscle. With the enactment of 21st Century Cures Act, the FDA biomarker qualification process has become more streamlined. In addition to developing
a clear framework for regulatory qualification, there is optimism towards expediting the qualification process so that promising therapeutics can be safely tested in the clinic and ultimately provide patients access to new medicines.

Conflict of interest---- The authors declare that there is no conflict of interest.

REFERENCES

FDA (2008): Qualification Letter for Seven Biomarkers of Drug-Induced Nephrotoxicity in Rats.
FDA (2012): Qualification Letter for Seven Biomarkers of Drug-Induced Nephrotoxicity in Rats.
FDA (2015): FDA Letter of Support for Serum and Plasma Biomarkers: Myosin Light Chain 3 (My3), Skeletal Muscle Troponin I (sTNI), Fatty Acid Binding Protein 3 (FABP3), Creatine Kinase, Muscle Type (CK-M, the Homodimer CK-MM).
FDA (2016b): FDA Case Study: A Collaborative Effort to Qualify a Novel Biomarker.
Guflroy, M., Falahatpisheh, H., Biddle, K., Keeger, J., Obert, L., Walters, K., Goldstein, R., Boucher, G., Coskran, T., Reagan,
Safety biomarker applications in drug development


