Preface

Some of the challenges the pharmaceutical industry faces in the current research and development processes are: (1) a lengthy process that takes, on average, approximately 11 years from target identification to the development of a new medicine; (2) an ever increasing costly process; (3) an inefficient process where too many drugs fail before they reach the market because of a lack of efficacy or unacceptable toxicity, as well as postmarketing withdrawal due to rare serious adverse events; (4) drug--drug interactions or toxicity is not uncommon; (5) the increasing difficulty in identifying novel drug targets; and (6) the mode of action for many compounds is often unknown. This is a depressing reality. Where are the improvements in both quality and efficiency often claimed in the drug development process? Why haven't the advances in science and technology made a greater impact? How can improvements in the process reduce the already high cost of drug development generally?

To address some of these issues, the pharmaceutical industry is actively exploring the relationships between human genetics and drug responsiveness, susceptibility to disease, and disease severity. While research approaches and emphases may vary from company to company, the overarching goal of the industry is largely consistent: to discover and develop new medicines based on an improved understanding of patient response to drugs (positive or negative) and of diseases etiology. Pharmacogenomic (PGx) methods are aimed at determining the contribution of genetic differences in ADME, drug target, and disease genes to drug response, thereby improving the safety and efficacy of drug therapy through use of genetically guided treatments, an approach called *personalized medicine. Personalized medicine* is both one of the newest disciplines of medicine currently being used and very much an ongoing work in progress. Many drug companies have incorporated genetic research, such as the collection of genetic samples, into their drug development programs. And while only a few examples of true success stories have emerged during the past few years of research, it is clear that the current landscape is driving us toward a more widespread acceptance of personalized medicine. Currently many questions arise regarding the appropriate implementation of this technology: how can the industry go about delivering true business value while recognizing that the ability to address patients' demand for safer and more efficient novel drugs might be met by engaging this technology more fully. There are

concerns within the pharmaceutical industry about generating data that might be difficult to interpret in a regulated environment. There is also a growing appreciation for the challenges in translating this new information into clinical utility, including scientific, commercial, ethical, and policy challenges.

Pharmacogenomics and Personalized Medicine, which is part of the *Methods in Pharmacology and Toxicology* series, comprises chapters on selected aspects of pharmacogenomics and personalized medicine. Our overall intent is to assist both novice and experienced investigators in understanding the current scientific challenges in applying PGx to discovery and clinical development and in making appropriate decisions to engage in and interpret PGx research. Designed to share the experiences of leading experts in the field, the book is a useful guide for conducting PGx research--from discovery to the market, but we also aim to present a realistic perspective on the challenges, practicalities, and obstacles in applying pharmacogenomics. Generally, the book avoids statements such as "Pharmacogenomics is going to revolutionize the practice of medicine," which are neither realistic nor particularly useful to anyone.

The book presents an industry perspective on the implementation of PGx in research and development, in drug discovery, and in clinical trials, including recommendations for a systematic approach for assessing the feasibility and added value of PGx studies in clinical trials. It also provides guidance on the key logistical issues required to prepare the pharmacogenomics protocol and an informed consent form for sample collection and analysis, the strategies and resources for SNP marker selection, and genotyping in genetic association studies, and the study design and statistical methodologies for data analysis in PGx research. We have included an interesting view of the effect of genetic variation, as well as a description of recent PGx applications in drug metabolism, adverse drug reactions, and in a few selected therapeutic areas (epilepsy, Alzheimer's disease, psychiatry, oncology, HIV, cardiovascular diseases). Additional key topics,s such as the current regulatory environment and drug label implications, biomarker qualification and trial design, the co-development of drugs and diagnostics, and the translation of genomics biomarkers into clinical utility, are also covered. Furthermore, two chapters describe the current state of knowledge of PGx in rare and monogenic disorders and in children, which are currently less well covered in the published literature but deserve attention.

Pharmacogenomics and Personalized Medicine focuses on DNA data and associated analytical methodologies that are currently the more mature components of the evolving constellation of genomic sciences. However, complementary RNA-based studies are also being considered in some chapters. It is important to also acknowledge that remarkable progress is being made in complementary methodological areas such as *proteomics*, *metabolomics,* and *imaging*. Given the layered complexities of biological regulation, it is likely that reliable markers will be hybrids that will cross methodological disciplines. A program of persistent innovation is being required from the industry to balance near-term profit with the need to accommodate the increasingly competitive and changing landscape. Education and cooperation among experts from the scientific community, industry, and government are recognized as integral to greater success in personalized medicine. It is my hope

that the knowledge we share here regarding DNA information may be leveraged to create a useful foundation for further progress in personalized medicine, using other approaches that will benefit the pharmaceutical industry overall, and most importantly the patients.

Experts from the pharmaceutical industry, scientific community, and government have been invited to contribute their experience to this book. I would like to express my gratitude to all contributors for their enthusiasm to this work. Without their time and energy, *Pharmacogenomics and Personalized Medicine* would not have been possible.

Nadine Cohen, PhD

Contents

Chapter 2
Implementation of Pharmacogenomic Sample Collection in Clinical Trials

Deborah Sokol Ricci and Monique Franc

Contents

Abstract This chapter is intended to provide an overview of the operational considerations and potential obstacles that can be anticipated during the implementation of pharmacogenomic research in clinical trials. Particular attention is given to the elements of the protocol and of the informed consent and the considerations for collection of different sample types on a global level. The goal is to provide the reader with an appreciation for the study design elements on an operational level rather than on a scientific or statistical study design level. Educational efforts by various working groups to harmonize global standards are also outlined and will provide the reader with an overview of the ongoing efforts to promote global genomic research in the present day.

Keywords biomarker, genomics, pharmacogenomics, global sample handling, genetics, global regulations, local regulations, sample coding, exploratory research

Deborah Sokol Ricci
Johnson and Johnson Pharmaceutical Research and Development, L.L.C., 1000 Route 202, Raritan, NJ 08869
dricci@prdus.jnj.com

N. Cohen (ed.) *Pharmacogenomics and Personalized Medicine*,

Glossary

AAPS Pharmacogenetics and Pharmacogenomics Focus Group	A focus group working towards information exchange for developments in pharmacogenomics and pharmacogenetics.
Adverse event	In pharmacology, any unexpected or dangerous reaction to a drug.
Anonymization	Samples are double coded and labeled with a unique second number. The link between the clinical study subject number and the unique second number is deleted.
Assay validation	Optimization of an assay protocol with respect to sensitivity, dynamic range, signal intensity, and stability.
Comparative genome hybridization (CGH)	Comparative genomic hybridization (CGH) measures DNA copy number differences between a reference genome and a sample genome.
Complementary DNA (cDNA) library	A collection of cDNAs, each of which has been inserted in a DNA vector (e.g., a circular DNA plasmid) and replicated in a bacterium such as *E. coli*.
CPT tubes	Cell preparation tubes with sodium heparin, utilized for the separation of mononuclear cells from whole blood.
De-identification	Samples are double coded and labeled with a unique second number. The link between the clinical study subject number and the unique second number is maintained, but unknown to investigators and patients.
Deoxyribonucleic acid (DNA)	A molecule that encodes genetic information.
EDTA	A crystalline acid, $C_{10}H_{16}N_2O_8$, that acts as a strong chelating agent.
European Federation of Pharmaceutical Industries and Associations (EFPIA)	A focus group with a pharmacogenomics task force; this group has overlap with the Pharmacogenetics Working Group.
Exploratory analyses	General exploratory or research information collected from studies such as broad gene expression screening, whereby the markers studied have not reached the status of a probable valid biomarker.
Formalin fixation	Tissue fixation in a solution containing formalin.
Formalin fixed paraffin embedded (FFPE)	A method of preserving tumor tissue for pathological and other analyses.
Laser capture microdissection (LCM)	A method to collect a specific subset of cells from a slice of tumor tissue captured on a slide.
Loss of heterozogosity (LOH)	The loss of one parent's contribution to part of the cell's genome.
Paraffin embedding	A method of preserving fixed tissue (see formalin fixation).
PAXgene™	The PAXgene™ Blood RNA System consolidates and integrates the key steps of whole blood collection, nucleic acid stabilization, and RNA purification. By minimizing the unpredictability associated with RNA processing, the system provides enhanced accuracy of intracellular RNA analysis.
Pharmacogenetics for Every Nation Initiative (PGENI)	An initiative to enhance the understanding of pharmacogenetics in the developing world.
Pharmacogenetics Research Network	Enables a network of multidisciplinary research groups to conduct studies addressing research questions in pharmacogenetics and pharmacogenomics with a goal to populate a knowledge base with data.

(continued)

Glossary (continued)

Pharmacogenetics Working Group	A voluntary and informal association of pharmaceutical companies engaged in research in the science of phamacogenetics.
Pharmacogenomics	The study of how variations in the human genome affect the response to medications.
Polymerase chain reaction (PCR)	A method to enzymatically replicate DNA.
Definitive analysis	Preplanned and prespecified research analyses.
Protein	Proteins are fundamental components of all living cells and include many substances, such as enzymes, hormones, and antibodies, that are necessary for the proper functioning of an organism.
Real time quantitative PCR (qRT-PCR)	A method to quantify low abundance messenger RNA (mRNA), enabling a researcher to quantify relative gene expression at a particular time, or in a particular cell or tissue type.
Regulations	A legal restriction promulgated by government administrative agencies through rulemaking supported by a threat of sanction or a fine.
Ribonucleic acid (RNA)	RNA serves as the template for the translation of genes into proteins.
RNALater	A reagent used to immediately stabilize RNA from tissues.
Sample coding	A method to label samples whereby personal identifiers are not present.
SELDI-TOF	Surface Enhanced Laser Desorption/Ionisation Time of Flight mass spectrometry. A methodology utilized for proteomic analyses.
The Council for International Organizations in Medical Sciences Pharmacogenetics Working Group.	A working group formed to consider issues related to pharmacogenetics with respect to drug development and regulatory, ethical, educational, and economic issues.

1 Introduction

The value of pharmacogenomics in clinical trials has become increasingly recognized, not only by the pharmaceutical industry, but also by regulatory agencies (as evidenced by a growing regulatory framework for pharmacogenomic research) *(1–3)* and by the general public (as evidenced by the increasing media attention on "personalized medicine") *(4)*. In a relatively short time, there has been significant evolution in the acceptance of this science as both a supplement and in some cases an alternative to the classical paradigm of the drug development process; and this despite a relative paucity of guidelines, regulations, and global harmonization. Nevertheless, with the appropriate procedures, it is now possible to routinely collect samples for pharmacogenomic applications in clinical trials, although apprehensions and hurdles still linger in some locales. On an international level, it is predicted that genomic sample collection will become more mainstream in some jurisdictions and more cumbersome in others as knowledge, experience, and familiarity increase and as laws and regulations

are implemented and revised. Current efforts are ongoing to educate various bodies toward harmonization and standardization of regulations and practices applicable to pharmacogenomic collections globally. However, it is anticipated that harmonization will not happen in the near future and that a lot of work will be required to bring consensus across the nations.

It has become evident that there is no one-size-fits-all solution for implementing pharmacogenomic sample collections when operating in a global environment. Not only do requirements vary from country to country, but they frequently vary on state, provincial, and local levels. This requires a great degree of flexibility, since adjustments are often necessary at individual sites. However, the considerations described herein should allow the researcher to efficiently implement the collection of pharmacogenomic samples in clinical trials on an international level while maximizing the use and value of the collected samples and maintaining stringent standards for subject privacy.

The term "genomic" used throughout refers to both DNA and RNA. It can be debated whether RNA deserves the same level of stringency for sample and data handling as does DNA, or for that matter, whether either of these molecular endpoints should be handled any differently than any other biochemical or clinical endpoints (i.e., "genetic exceptionalism"). However, the reality today is that these endpoints *are* viewed as "special" and *do* receive unique attention by regulatory agencies. There currently are overarching expectations for a higher level of stringency for maintaining confidentiality and preventing unintended access to or release of genomic samples and data. Since both DNA and RNA are captured together in the recent flood of pharmacogenomic guidelines released by regulatory agencies (U.S. Food and Drug Administration [FDA], European Medicines Agency [EMEA], and Japan Ministry of Health, Labor, and Welfare [MHLW]), both molecular endpoints are treated in tandem in this chapter; although it is acknowledged that privacy and confidentiality expectations may not always be as strict for RNA-based research as they are for DNA-based research. Because proteomic analyses often accompany genomic-based research, methods to increase sample integrity are mentioned. However, since regulations around these analyses are not as stringently controlled, they are not addressed herein. A glossary of useful terms is provided on page 28.

2 Protocol

As is true with all clinical trial procedures, pharmacogenomic research on human subjects must be described in a protocol, whether harmonized within the clinical protocol itself or as a stand-alone document. In developing protocols for pharmacogenomic studies, there are a number of practical study design elements that are dictated by the goals of the trial and of the pharmacogenomic study, notably: i) whether participation will be optional or mandatory, ii) whether genomic endpoints are already known or will only be defined over the course of the trial according to emerging clinical issues, and iii) whether samples will be analyzed within the

context of the trial only or will also be retained for future research. Each of these parameters is associated with varying degrees of operational, ethical, and regulatory implications described below.

2.1 Optional Versus Mandatory Subject Participation

Most commonly, pharmacogenomics is integrated as an "add-on" component to the clinical trial, although increasingly, pharmacogenomic parameters are being factored into trial design. The pharmacogenomic component usually does not have a direct impact on patient medical care. It is, however, accompanied by sensitivities around privacy and the potential for discrimination. For these reasons, a decision must be made whether the provision of samples for genomic research is required of subjects enrolling in a clinical trial or whether it is sufficient to offer this to subjects as an optional component. When referring to subject participation, the term "optional" should be distinguished from "voluntary," since in a legitimate clinical trial, by definition, every procedure is "voluntary"—meaning that a person cannot be forced to undergo any procedure against his/her will (or that of his/her legally acceptable representative, as applicable). This is true for both optional and mandatory components of the trial. However, for optional procedures, refusal to consent would not compromise eligibility for the trial, whereas for mandatory procedures, refusal to consent would result in ineligibility for the trial.

Optional Participation

Optional subject participation in pharmacogenomic research is currently the most common and straightforward pharmacogenomic study design option. It generally is appropriate when there are no definitive genomic analyses to be performed, or when pharmacogenomic results are not critical to the outcome or design of a study, or when the proposed analyses are purely exploratory. Operationally, optional participation has no impact on the rate of enrollment or on the duration of the screening period. The rate of participation will largely be dependent on i) whether subjects are healthy volunteers or diseased patients, ii) the specific disease under investigation, iii) the delivery of the informed consent process, iv) the geographical location, and v) the specific patient population (e.g., pediatric). Although there are exceptions, agreement to participate in genomic research is often higher in healthy volunteers than in diseased patients, presumably since subjects have a somewhat different focus and motivation for their involvement in the trial. With the appropriate informed consent process, a participation rate for optional pharmacogenomic research of 80% or greater should be readily achievable in studies involving healthy volunteers. Participation rates tend to be lower in diseased patients, although not necessarily for all disease indications. The attitude of the investigator administering the informed consent also can have an appreciable influence on participation rates, owing to the impressionability of subjects in the face of perceived authority figures. The opinion

of one or a few other participants can similarly have a significant impact on participation rates, particularly in the case of group consent procedures owing to the groupthink phenomenon of human behavior. Acceptance may be dictated by cultural background, personal experience, and level of education of the subjects. Education of investigators, study coordinators, regional monitors, and local trial managers is of the utmost importance in maximizing participation rates while simultaneously avoiding the coercion of subjects. The degree of pushback from ethics committees on optional participation is minor except in occasional cases, such as when the term "optional" is deemed to imply "unimportant" and therefore "unnecessary;" but this can usually be readily resolved through education on the value of genomic sample collections in clinical trials. The obvious and most important limitation of optional participation is that it is possible that not all subjects enrolled in the trial will agree to provide genomic samples, and thus any unusual clinical outcome occurring in non-consenting individuals (e.g., adverse events) could not be investigated using a pharmacogenomic approach (unless the subjects were approached again retrospectively for their consent). However, in the absence of specific analyses that are critical to the success of the trial, the limitation of having samples from less than 100% of the subjects is generally outweighed by the operational and ethical impediments that can be expected by imposing mandatory participation.

Mandatory Participation

Mandatory participation may be necessary in some cases, notably when genomic results are i) used to determine eligibility for the trial (i.e., inclusion/exclusion criterion; e.g., exclusion of CYP2D6 poor metabolizers), ii) are critical to the successful analysis of the clinical data (e.g., EGFR stratification), or iii) are requested by regulatory authorities (e.g., valid biomarkers known to be relevant to the compound under investigation). Mandatory participation may pose a number of operational challenges. Site selection will be affected, since some countries do not permit collection or export of genomic samples. Therefore, mandatory participation would preclude the execution of clinical trials in certain countries, and consequently in corresponding ethnic groups. Even within countries that authorize genomic sample collection, local regulations may preclude it, and therefore necessary assurances that genomic sample collection is permissible should be sought before sites are selected. Ethics committees will generally approve the mandatory requirement if the rationale is explicit and justified; they generally will not approve it in the absence of definitive analyses (see below). Subject enrollment rates may be compromised to varying degrees, since some prospective subjects may be uncomfortable with the idea of genomic research being conducted on their samples. Importantly, these subjects would be denied access to the trial and to the potential benefits of treatment with a novel drug; however this may not be an issue if alternative therapies are available. Subjects should be clearly informed that refusal to consent or subsequent withdrawal of consent to conduct research on their genomic sample would result in ineligibility for continued participation in the trial, although ethical decisions may override this requirement. Where genetic data

are intended to be used to determine eligibility, the impact on the duration of the screening period should be anticipated. This will be dictated primarily by the prevalence (frequency) of the genomic result of interest, the turnover time for generating the data, and the overall sample size. The advantages of mandatory participation are that samples are available for each and every subject (assuming no samples are lost due to mishandling), and that studies can be designed and optimized based on genomic information.

2.2 Definitive Versus Tentative Analyses

Definitive analyses refer to preplanned and prespecified genomic endpoints that are committed to being analyzed in a clinical trial. Tentative analyses refer to analyses that may be performed only as necessary, if it is hypothesized that this might help to resolve unanticipated issues with the clinical data. The terminology "definitive analyses" is used here preferentially over "prospective analyses" or "hypothesis testing," which can have statistical or trial design connotations.

In principle, a clinical protocol should provide a thorough description of all endpoints to be measured in a clinical trial. For definitive analyses, this may consist of specific genes, genetic loci, or transcripts, whether few or many. However, by its inherent nature, pharmacogenomics is frequently utilized to help address unexpected clinical results (e.g., pharmacokinetic outliers, variable efficacy, adverse events). Therefore, room must be made to accommodate this valuable application by allowing for tentative analyses in the absence of a specific preexisting hypothesis. There are several points for consideration when collecting samples, even if only for tentative analyses. The operational aspects include: establishment of contracts, coordination with central or local laboratories, logistics for shipment, education of investigators and site staff, sample and consent tracking, additional informed consent procedures, and time for collection of the additional specimen(s). The cost associated with the collection, handling, and potential long-term storage of the sample must also be taken into consideration, but is generally nominal compared to retrospective sample collections. For example, the cost of a DNA collection (from whole blood), including disposables (e.g., collection tubes/kits), phlebotomy, sample shipping and handling, DNA extraction, and sample and data management, can currently be estimated at approximately US$50–100 per sample, excluding costs associated with long-term sample management. Costs are higher for DNA obtained from various tissues (e.g., tumors), because of the more involved process of sample acquisition and preparation. Similar procedures for RNA extraction are also more costly. It may be challenging to justify this cost, particularly in very large Phase III or Phase IV trials, in the absence of a definitive analysis to be performed. However, this must be weighed against the risk and cost associated with not having the samples in the event of an emerging issue, as well as the lost opportunity to conduct large-scale pharmacogenomic research on banked samples (e.g., genomic studies of disease, rare adverse events, drug class effects). Retrospective collection of DNA from subjects may be an option,

but is exceedingly more costly (as high as tenfold) and more logistically challenging than is prospective DNA collection, and is associated with potentially significant delays in attempts to address emerging issues. Ethics committee resistance may sometimes be encountered if the collection for tentative analysis is not viewed as being critical or necessary for the success of the trial, since there is reluctance to collect human biological samples if they may sit endlessly in a freezer, never to be analyzed. Increasingly, however, ethics committees are conscious of the "insurance value" of precautionary genomic sample collection and will normally approve collection for the purpose of tackling unanticipated issues, if this rationale is clearly stated. The justification is further reinforced if samples are also intended for storage for future research. Since open-ended proposals for research on genomic samples is generally viewed unfavorably by ethics committees, the listing of candidate genes/loci/transcripts that may potentially be analyzed will ordinarily satisfy committee needs for a definition of the scope and boundaries of the possible use of the samples. This list will ordinarily include genomic endpoints relevant to pharmacokinetics, potential adverse events, mode of action, and the disease under investigation (as appropriate). To accommodate situations in which ethics committees do not approve genomic sample collections, it is advisable to state explicitly in the protocol that the clinical protocol can be approved independently of the pharmacogenomic component. The qualifier "where local regulations permit" throughout is useful for this purpose and avoids the need for protocol amendments.

2.3 Sample Banking for Future Research

Since the value of genomic samples generally increases as sample sizes increase, it is recommended that the protocol be conceived to allow for long-term sample storage for the purpose of future research in order to maximize the potential value of the samples, particularly in clinical trial settings where clinical data collection is standardized and of high quality. As scientific discoveries are made and as science evolves, valuable research can be done in the future on samples collected today. Some study participants may not be comfortable with long-term storage of their samples, particularly if future uses are unknown. Therefore, processes that introduce two levels of participation, i.e., one for research specifically related to the trial (including both definitive and tentative analyses) and another for storage of samples for future research of broader scope, offer a higher degree of flexibility and help to maximize subject participation rates. Although most ethics committees are agreeable to the banking of samples for future research, many will require the delineation of boundaries and limitations for the scope of the research that may be conducted on the samples. For instance, research may be limited to that which is relevant to the drug or drug class and/or the disease or therapeutic area under investigation. Many ethics committees will permit the indefinite storage of samples; some insist on sample destruction after a predefined storage period; and others will not approve of this application at all. Processes should therefore allow for tracking timelines for sample destruction.

Ethics committees and subjects will generally be amenable to long-term sample storage for future research, provided that there are sufficient assurances that stringent processes and standards for patient privacy/confidentiality are in place. Patient privacy can be achieved by measures that minimize the possibility of linking genetic data back to a subject's identity. This can be accomplished by a number of methods, including i) de-identification of samples such that a coded sample is relabeled with a unique second code, while maintaining a link between the two codes (i.e., double-coded); or ii) anonymization of the samples such that the link between the two codes of a double-coded sample is permanently deleted. Anonymization offers the maximum achievable level of security, while still allowing for genotype-to-phenotype correlative analyses to be undertaken. The deletion of the coding key linking the sample(s) to the subject's study identifier provides an additional level of security over de-identified data, as it renders obsolete the coding key used for the re-identification of subjects via their original subject identifier. The purpose of anonymization is to express the deliberate intent to not re-identify subjects. This is in contrast to de-identification which maintains the intent to link back to the subject identifier, if necessary. Consequently, actions such as returning results, sample withdrawal, clinical monitoring, or patient follow-up cannot be undertaken on anonymized samples. A common misconception is that anonymization severs the link between the sample and the corresponding clinical data for a given subject; when, in fact, what is lost is the ability to link the new subject identifier to the original subject identifier. Anonymization does not interfere in any way with relating genotype data to phenotype data, since genomic samples and data are coupled to the clinical data prior to anonymization. Consequently, it is critical that all relevant clinical data be fed into the anonymization procedure since, by definition, anonymization is a permanent, irreversible process that does not allow retrospective addition of data. It is not unusual to find that the level of participation in sample storage for future research is generally slightly lower than for research that is directly relevant to the trial, although this would depend on the assurances offered for protection of patient confidentiality. Interestingly, some subjects will choose to participate only in the storage for future research component because of the fact that the risk of linking genetic data back to their identity is lessened through the anonymization/de-identification process (as applicable).

It is recognized that anonymization can be susceptible to reconstruction of the link between anonymized genetic data and a study subject identifier by means of comparing an anonymized dataset with a separate dataset that contains the subject identifier. The reason for this is that the clinical data set can serve as a "clinical barcode" or "clinical fingerprint" that uniquely identifies a subject. For this reason, the term "*not possible*" is not accurate as it relates to the ability or possibility to link anonymized samples/data back to a subject *(5)*. Anonymization should therefore always be accompanied by specific policies or standard operating procedures (SOPs) prohibiting reconstruction of any kind of link between genetic data and the original study subject identifier. In addition, access to datasets ideally should be restricted, and no one individual should have access to all information necessary to re-establish a patient's identity. It should be noted that anonymization may have regulatory consequences, because currently, data generated from anonymized samples may not always

be used for regulatory decision-making. It should also be noted that whereas some countries require anonymization of samples prior to storage, others do not permit anonymization. Therefore, a means of tracking country of origin is essential.

3 Informed Consent

With few exceptions, obtaining legally effective, voluntary informed consent is a fundamental prerequisite for conducting research on human beings. Surprisingly, despite a long history of medical research, there is no single universally accepted list of basic elements of informed consent, although the International Conference on Harmonization (ICH) has significantly improved harmonization of informed consent requirements on a global level. The policies and regulations which allow informed consent to be legally effective vary on national, state, and local levels, which poses a challenge to clinical research that is conducted on an international level as is commonly encountered in clinical trials *(6)*. Institutional review boards (IRBs) and independent ethics committees (IECs) serve as gatekeepers to ensure compliance with requirements and regulations.

It is important to appreciate that informed consent is a process, not just a form for the subject to sign. A key requirement for informed consent is that the information be presented in a manner that is understood by the prospective subject (or legally acceptable representative), and that it enable the individual to voluntarily decide whether or not to participate. Regrettably, informed consent forms are more often geared toward legal protection of the investigator and study sponsor than toward providing information to the subject in a manner that is truly understandable, educational, and meaningful. In the absence of a test to ascertain the true degree of understanding by the subject, every effort must be made to simplify the informed consent process and to strike a balance between providing sufficient information for a subject to make a reasoned decision about whether to participate while simultaneously protecting the legal interests of those conducting the study. The specific elements of consent and the verbiage selected are equally important.

3.1 Readability and Understanding

There is currently no universal standard for assessing how much information is understood or retained by a prospective study participant *(7, 8)*. Readability algorithms such as Flesch Reading Ease *(9, 10)* and Flesch-Kincaid Grade Level, which are based on scientific linguistics that calculate average sentence length and number of syllables per word to generate an index of difficulty, may prove valuable in the design of the informed consent form. However, readability should not be confused with understandability and there is some debate about whether readability statistics do indeed result in improved understandability and retention *(11–13)*. Oversimplification of sentences and words can reach a point of diminishing return

and actually make understanding more difficult, in part by unnecessarily lengthening the overall text *(14, 15)*. However, a combination of techniques, including the use of short sentences, monosyllabic words, simple phrases, active voice, paragraphs no longer than four to five sentences, sentence structure in subject-predicate position, use of ample white space between paragraphs, left justification with right ragged margins, minimum 10-point font size, avoidance of nouns created from verbs, avoidance of multiple negatives, and limitations on the total amount of information provided, can work synergistically in creating an informed consent document that is understandable by the average subject. By whatever means achieved, a readability level approximately equivalent to that of a 12-year-old child, or comparable to that of a typical newspaper, would be suitable for the average layperson (16). The goal is to communicate the information in a thorough, clear, and concise way, while avoiding information overload.

Owing to the sensitivities surrounding genomic research (whether perceived or real), and the corresponding unique considerations applicable to genomic samples and data that are generally not necessary for other samples types (e.g., implications for family members due to the heritability of DNA), it is not uncommon for separate informed consent forms, one for the clinical trial and one for genomic research, to be used. This allows for the general details of the trial to be more effectively communicated without distraction by the specific details and sensitivity issues that are associated with the genomic samples. A subject might otherwise be overwhelmed trying to decipher which conditions apply to which samples in which parts of the trial. The dual consent format can improve readability and understanding by allowing the subject to consider the issues presented in each consent form separately. Since participation in genomic research is more commonly offered as optional, and since local regulations may preclude genomic sample collection at some investigational sites, the dual consent model allows a subject or site to readily opt out of genomic research. This model also helps to make it clear that the pharmacogenomic component is a separate substudy and that agreement or refusal to participate is unrelated to eligibility for the trial (as applicable). For clinical trials involving mandatory participation in genomic research, it would be reasonable to merge the genomic ICF with the clinical ICF; however, for ease of readability, the details pertaining specifically to the genomic component would best be contained within one section rather than interwoven with the main clinical ICF.

3.2 Elements of Informed Consent

In the United States, the Code of Federal Regulations (CFR) 21 CFR 50.25 (2007) lists eight basic elements of informed consent, and six additional elements to be included where applicable (17). In contrast, the Council for International Organizations of Medical Sciences (CIOMS) Guideline 5 recommends 26 basic elements, many but not all of which are an extension of the eight basic elements of informed consent from the CFR (see Table 1) (18). To date, there are no internationally recognized regulations that dictate the basic elements of informed consent for

Table 2.1 CIOMS essential information for prospective research subjects (Guideline 5 of the International Ethical Guidelines for Biomedical Research Involving Human Subjects (18))

Before requesting an individual's consent to participate in research, the investigator must provide the following information, in language or another form of communication that the individual can understand:

1. that the individual is invited to participate in research, the reasons for considering the individual suitable for the research, and that participation is voluntary;
2. that the individual is free to refuse to participate and will be free to withdraw from the research at any time without penalty or loss of benefits to which he or she would otherwise be entitled;
3. the purpose of the research, the procedures to be carried out by the investigator and the subject, and an explanation of how the research differs from routine medical care;
4. for controlled trials, an explanation of features of the research design (e.g., randomization, double-blinding), and that the subject will not be told of the assigned treatment until the study has been completed and the blind has been broken;
5. the expected duration of the individual's participation (including number and duration of visits to the research centre and the total time involved) and the possibility of early termination of the trial or of the individual's participation in it;
6. whether money or other forms of material goods will be provided in return for the individual's participation and, if so, the kind and amount;
7. that, after the completion of the study, subjects will be informed of the findings of the research in general, and individual subjects will be informed of any finding that relates to their particular health status;
8. that subjects have the right of access to their data on demand, even if these data lack immediate clinical utility (unless the ethical review committee has approved temporary or permanent nondisclosure of data, in which case the subject should be informed of, and given, the reasons for such nondisclosure);
9. any foreseeable risks, pain or discomfort, or inconvenience to the individual (or others) associated with participation in the research, including risks to the health or well-being of a subject's spouse or partner;
10. the direct benefits, if any, expected to result to subjects from participating in the research;
11. the expected benefits of the research to the community or to society at large, or contributions to scientific knowledge;
12. whether, when, and how any products or interventions proven by the research to be safe and effective will be made available to subjects after they have completed their participation in the research, and whether they will be expected to pay for them;
13. any currently available alternative interventions or courses of treatment;
14. the provisions that will be made to ensure respect for the privacy of subjects and for the confidentiality of records in which subjects are identified;
15. the limits, legal or other, to the investigators' ability to safeguard confidentiality, and the possible consequences of breaches of confidentiality;
16. policy with regard to the use of results of genetic tests and familial genetic information, and the precautions in place to prevent disclosure of the results of a subject's genetic tests to immediate family relatives or to others (e.g., insurance companies or employers) without the consent of the subject;
17. the sponsors of the research, the institutional affiliation of the investigators, and the nature and sources of funding for the research;
18. the possible research uses, direct or secondary, of the subject's medical records and of biological specimens taken in the course of clinical care (see also Guidelines 4 and 18 Commentaries);
19. whether it is planned that biological specimens collected in the research will be destroyed at its conclusion, and, if not, details about their storage (where, how, for how long, and final disposition) and possible future use, and that subjects have the right to decide about such future use, to refuse storage, and to have the material destroyed (see Guideline 4 Commentary);

(continued)

Table 2.1 (continued)

20. whether commercial products may be developed from biological specimens, and whether the participant will receive monetary or other benefits from the development of such products;
21. whether the investigator is serving only as an investigator or as both investigator and the subject's physician;
22. the extent of the investigator's responsibility to provide medical services to the participant;
23. that treatment will be provided free of charge for specified types of research-related injury or for complications associated with the research, the nature and duration of such care, the name of the organization or individual that will provide the treatment, and whether there is any uncertainty regarding funding of such treatment;
24. in what way, and by what organization, the subject or the subject's family or dependants will be compensated for disability or death resulting from such injury (or, when indicated, that there are no plans to provide such compensation);
25. whether or not, in the country in which the prospective subject is invited to participate in research, the right to compensation is legally guaranteed;
26. that an ethical review committee has approved or cleared the research protocol.

genomic research. The industry's Pharmacogenetics Working Group has prepared an elegant and comprehensive compilation of elements of informed consent for consideration in pharmacogenomic research studies (19), the essence of which is captured below, in addition to that of the authors' experience. The specific consent elements selected for a particular pharmacogenomic study and the details thereof will generally be a reflection of the policies or standard operating procedures of the sponsor, the specific trial design, the alignment with local laws and regulations, and the concessions made for readability and understanding.

It is beyond the scope of this chapter to evaluate the validity of concerns held by study participants, government bodies, ethical review boards, or investigators regarding risks or potential risks associated with the generation, use, and disclosure of genomic data. The major concerns undoubtedly stem from the heritable nature of DNA, the potential misuse and misinterpretation of genomic data, the shortage of policies and laws regarding the use and misuse of genetic information, and the fear of potential stigmatization and discrimination (20). Additionally, the banking of samples for future use of potentially unknown scope also raises some legitimate concerns for obtaining truly valid informed consent *(21, 22)*. In response to these concerns, activities related to genomic research are generally conducted under a higher level of stringency with regard to privacy protection and confidentiality.

The eight elements of informed consent of the U.S. Code of Federal Regulations (CFR) are presented below and are used as a framework for additional considerations for informed consent for pharmacogenomic research. The 26 basic elements of consent from CIOMS are presented in Table 2.1. A balance must be sought between the number of elements to include in order to adequately inform the subject and to comply with local regulations, while maintaining an acceptable level of readability and avoiding information overload.

CFR Element 1: A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental.

Specific considerations for pharmacogenomic studies:

A definition of "DNA" and "genes" and the heritable nature thereof should be presented in simple terms. For RNA-based studies, the nonheritable nature of RNA should be stated, although the direct derivation of RNA from DNA should be acknowledged.

A statement explaining the reason that DNA/RNA is being collected and how this will bring value to the clinical trial or to science in general, both in the short term and in the long term, should be included. In the case of definitive analyses, some key highlights on the relevance of specific endpoints that will be analyzed should be provided. If the samples are only being collected for tentative analyses, subjects should not leave with the impression that their samples will definitely be analyzed. Some subjects may feel that the additional sample collection is not justified and may opt not to participate on these grounds. It should be clear that the results from the research will not be used in the subject's medical care (where applicable). In the case of prescreening for trial eligibility, subjects should be informed that results of the genetic tests will be used to determine eligibility for the trial.

Although not part of the classical definition of pharmacogenomics, disease-genetics studies may overlap with pharmacogenomic studies since genetic factors that determine disease etiology or subtype may influence response to drugs. It should therefore be clearly stated if research related to disease genes/loci will be conducted.

The procedure for collecting the genomic sample (e.g., blood draw, buccal swab, tumor or tissue biopsy), including the volume or size of the sample and the timing of sampling should be described. For RNA studies, it is often necessary to collect samples at multiple time points; therefore the number of samples, time points, and volume at each time point should be defined.

The scope of the intended use of the samples should be stated in order to define the boundaries of what can be done with the samples. This can range from the analysis of one or a few candidate genes/loci/transcripts that are related to the drug(s) or indication(s) under investigation to broader research that is not directly relevant to the trial. It is insufficient to state that "genetic research will be done on the samples." Some ethics committees may require that these specific endpoints be listed in the informed consent, whereas others take the position that gene names are not meaningful to the nonexpert. One compromise is to make available the list of genes only upon a subject's request. If large-scale or genome/transcriptome-wide investigations might be conducted, this should be stated. In such cases, it is generally sufficient to state that "thousands of genes/RNAs will be analyzed in relation to…." It can be useful in the case of commercially available analytical platforms to include the name of the platform and/or web link to allow easy access to the list of specific genes/transcripts that will be analyzed.

If the samples will be retained for research in the future, this must be clearly stated, including both known uses and potentially unknown uses (if applicable). There have been many ethical discussions about the validity of consent for future unknown research on human tissue and samples from DNA biobanks *(21, 22)*. Since there are still many unknowns associated with the information contained within the genetic code, subjects should be made aware that there is uncertainty about the information the samples could potentially yield in the future. Ideally and where possible, subjects should be given the opportunity to agree separately to the storage of samples for future research. The duration of sample retention, whether finite or indefinite, should be specified, as well as any possibility of the perpetuation of samples (e.g., whole-genome amplification, creation of immortalized cell lines, etc.).

A statement describing the degree to which the tests can or cannot be used to make a diagnosis or treat a person for a certain disease should be included. For research-grade tests that cannot be used to make any diagnosis, the term "DNA research" is recommended over "genetic testing" since the latter comes with a diagnostic connotation (e.g., "genetic testing" for cystic fibrosis).

If additional clinical information is to be collected exclusively for the purpose of the pharmacogenomic research component (e.g., ethnicity information, family history, etc.), this should be stated, and the additional information should only be collected from subjects who choose to participate in DNA research.

CFR Element 2: A description of any reasonably foreseeable risks or discomforts to the subject.

Specific considerations for pharmacogenomic studies:

Physical risks: The physical risks associated with genomic sample collection are generally minor, since sampling customarily involves a blood draw or buccal swab, but should still be stated. If a more invasive sampling technique is required (e.g., tumor or tissue biopsy) the physical risks must be disclosed. Often, pharmacogenomic samples can be collected in tandem with other sample collections in the trial. However, any added risks associated with the collection of additional sample volumes/amounts intended specifically for pharmacogenomic purposes (if any) should be stated.

Emotional, psychological, financial, and social risks: The greatest perceived risk of genomic research is that of the potential for misuse of genetic information consequent to intentional or unintentional disclosure to third parties or to the subjects themselves (or their relatives). For instance, genetic results revealing a higher risk for a certain disease for the subject or subject's family can potentially be worrisome to some participants and, in theory, could have implications for insurability, employability, or eligibility for adoption, among other things. These potential risks should be stated, but should also be represented realistically and not be an overexaggeration that causes unnecessary alarm, which could have an unfounded negative impact on the conduct of the intended pharmacogenomic studies. Where genomic analyses are not of diagnostic grade, the potential for discrimination from misuse of the data is greatly reduced. Exploratory pharmacogenomic research generally does not fall under the category of "genetic testing" as understood by insurance companies (with some exceptions). In such a case, subjects should be informed that they would not

need to inform insurance companies that they had previously undergone genetic testing, a parameter that is used in underwriting in some jurisdictions.

Ethics committee opinions vary considerably regarding the amount of detail to include under the risks section of the informed consent. However, as per the ICH *(23)* and the U.S. CFR *(17)*, stated risks should be "reasonably foreseeable" and not an exhaustive list of what in theory could happen, particularly when there is no concrete history or evidence of such risks. Nevertheless, risks exceeding those of everyday life must be included in the consent process *(17)*, although a statement that "the chance of this happening is very small," may be appropriate.

CFR Element 3: A description of any benefits to the subject or to others which may reasonably be expected from the research.

Specific considerations for pharmacogenomic studies:

Pharmacogenomic studies generally offer limited direct benefit to the sample donor since investigations i) are not customarily conducted as part of clinical care, ii) are generally exploratory in nature, and iii) are not intended for the purpose of making diagnoses (with some exceptions). It should therefore be explained in the consent process that the information from the pharmacogenomic study may benefit others in the future by leading to the discovery of safer and more effective drugs or better understanding of the disease. The sense of helping the population at large should be highlighted. It cannot be ignored that clinical designs, such as dose selection based on genotype, could potentially benefit the patient by optimizing the dose for that patient, or by excluding subjects who might otherwise suffer predictable adverse events (e.g., excluding CYP2C9 poor metabolizers in a warfarin drug-drug interaction study). It is also conceivable that subjects who request their data may also follow up on these results in an accredited diagnostic setting and eventually learn information that could benefit them in the future (e.g., knowing one's *CYP2C9* and *VKORC1* genotype in the event of future warfarin therapy). The use of diagnostic-caliber assays may also be of direct benefit to subjects. However, more commonly, analyses are conducted in research mode and the potential to benefit the subject is nominal. Rather, the likely benefit is to the scientific community or to the drug development process.

CFR Element 4: A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.

Specific considerations for pharmacogenomic studies:

Generally, for pharmacogenomic studies, the only alternative is not to participate. Where participation is mandatory, the alternative may be to participate in another trial for which there is no mandatory pharmacogenomic component.

CFR Element 5: A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained and that notes the possibility that the FDA may inspect the records.

Specific considerations for pharmacogenomic studies:

Information on sample coding and storage procedures should be described but should not be a lengthy discourse on processes (e.g., how a sample and corresponding data are anonymized). A description of the impact of these procedures on

patient privacy is more useful and should be the emphasis. For example, in the case of anonymization, it would be sufficient to explain to participants that their sample will be labeled with a new number that is not linked to their original study number, which makes it very difficult (but not impossible) to link their genomic sample and data back to them.

Examples of safeguards to prevent unauthorized access to or loss of the samples should be provided, e.g., building card-key access, locked freezers, etc. It is not advisable to simply state that samples and data will be maintained securely and that confidentiality will be maintained.

Since the value of genomic samples increases with the number of samples, it has become common for samples to be shared or pooled among research groups. Policies for the sale, loan, donation, or transfer of samples to third parties, including research partners, biobanks, service providers, and commercial entities, should be stated. The type of research that may be conducted by these parties should also be mentioned as well as the possibility that samples will be sent to countries where privacy regulations may not be as stringent (if applicable, see below).

The degree to which access to data will be safeguarded, including a list of parties who will or will not have access to the data and measures to control this access (e.g., secure databases, passwords, locked archives, policies, etc.) should be briefly stated.

A statement on publication and presentation of data, including the possibility of uploading genetic and clinical data into public databases, should be included.

The extent to which pharmacogenomic data and documentation will be segregated from medical records should be addressed.

CFR Element 6: For research involving more than minimal risk, an explanation as to whether any compensation and an explanation as to whether any medical treatments are available if injury occurs and, if so, what they consist of, or where further information may be obtained.

Specific considerations for pharmacogenomic studies:

Pharmacogenomic studies usually do not involve more than minimal physical risk, unless a tissue or biopsy sample is being collected exclusively for this purpose, in which case details of compensation or medical treatments in the event of injury should be stated. Otherwise, in keeping with improved readability and to avoid unnecessary repetition, it would be sufficient to explain in the pharmacogenomic informed consent that any medical injury sustained from the collection of the pharmacogenomic sample will be handled in the same way as described for the main clinical trial. Since pharmacogenomic research potentially could involve other non-physical risks, an explanation of what compensation would consist of in such a case, if any, should be provided.

CFR Element 7: An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of a research-related injury to the subject.

In general, there are no specific considerations for pharmacogenomic studies.

CFR Element 8: A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

Specific considerations for pharmacogenomic studies:

In addition to stating that participation is voluntary, it is also important to specify whether participation is optional or mandatory.

A description of the process for withdrawal of consent and implications for eligibility for the trial should be provided. Where participation is mandatory, it must be clear that withdrawal of consent from pharmacogenomic research would result in ineligibility for continued participation in the trial (if applicable).

The time frame during which a subject can withdraw from the pharmacogenomic component of the study should be defined, with mention of any conditions that would not allow sample destruction upon withdrawal from the trial (e.g., anonymization or pooling of samples or data). The available options for the fate of the samples upon withdrawal should be offered (e.g., samples will be maintained according to original consent, or samples will be destroyed).

An additional element of informed consent for pharmacogenomic studies that is commonly required by IECs and IRBs is a description of procedures for the return of research data to subjects and the implications thereof. The Pharmacogenetics Working Group has published detailed recommendations on the return of pharmacogenetic research data to research subjects *(24)*. In brief, some important considerations are i) the subjects' right to know and their right not to know; ii) the extent to which the data are interpretable and the ability to make a diagnosis; iii) the conditions under which the data will be generated (research versus diagnostic) and the consequent need to have tests repeated by an accredited diagnostic laboratory, noting that diagnostic-caliber tests may not be available for the endpoints analyzed; iv) the implications of knowing one's results (risks and benefits); v) the potential impact on family members should they learn of these results; and vi) access to genetic counseling. It is beyond the scope of this chapter to address each of these points. However, on a practical level, the overall approach that is currently most broadly accepted by ethics committees is to provide genomic research results to subjects only upon their explicit request, and this only after the limitations of the interpretability of the results have been clearly explained during the informed consent process. If results are to be returned to subjects, the conditions under which this would occur, who might see these data, the risk of the data turning up in subjects' medical files, and the degree of interpretability of the data and the potential impact on family members should be stated, including a statement about whether genetic counseling would be provided. In compliance with local regulations in some countries, ethics committees will not allow the return of genomic results to subjects unless the data were generated under conditions that allow clinically relevant interpretations to be drawn. In such cases, subjects should be informed upfront that they will not be entitled to receive their data, including the reason for this restriction.

The informed consent requirements in the U.S. CFR regulations are not intended to preempt any applicable federal, state, or local laws which require additional

information to be disclosed for informed consent to be legally effective *(17)*. Additional elements may include any of the 26 elements of informed consent from CIOMS or any other requirement that satisfies local laws and regulations *(25)*. Ethics committees are entitled to impose additional requirements to ensure compliance with institutional policy and local law.

The implementation of pharmacogenomic research in clinical trials does not end with the protocol and informed consent; additional documentation and infrastructure are required to successfully implement these studies in the context of clinical trials. Additional documentation minimally includes forms for withdrawal of consent, forms for coordinating data return to subjects, and templates for the accurate and consistent reporting of data. Also necessary are case report forms to capture consent status as well as information on local requirements (e.g., limitations on duration of storage) or any additional phenotypic data collected exclusively for the pharmacogenomic component. Obviously, SOPs supporting sample and data processing and handling, anonymization, and data return, and other relevant aspects, are recommended.

4 Considerations for Sample Collections

Because pharmacogenomic studies are being increasingly used to create and validate diagnostic and prognostic signatures and to support toxicological and functional studies that underlie the regulatory filings for new drug submissions, it is increasingly important to create standardized and robust methods for sample procurement and processing in addition to the parameters listed above. Correct sample identification requires error-free handling during all stages of sample collection and storage. Informatics systems should be specifically designed to register and track samples in addition to housing genomic and other data in a robust manner. All procedures related to sample acquisition intended for genomic research should be accompanied by guidelines or formal SOPs to ensure the quality and integrity of the samples and related data collected for these purposes. The following sections will outline processes that should ensure quality standards for sample procurement that will enable accurate and predictable data generation. Standardized procedures for sample collections during the conduct of clinical studies will significantly improve the reliability of the results obtained, while standardized isolation procedures for DNA, RNA, and protein will improve the overall quality of the results. The following sections describe multiple aspects to consider when collecting samples from multicenter clinical trials with the intent of conducting robust genomic and proteomic analyses worldwide.

4.1 Blood Samples

Blood samples are an excellent source of large amounts of DNA that can be used for genetic testing, either via candidate gene or genome-wide screens. In general,

whole blood is collected under standard conditions into vacutainer tubes containing EDTA and kept at room temperature or 4°C for overnight shipments for processing and/or storage. Alternatively, immediate processing and storage can be done at temperatures ranging from -20°C to -70°C, although this method is not convenient for multicenter clinical trials. DNA yields are optimal when whole blood samples are immediately processed to isolate DNA. Whole blood samples stored at ambient temperatures for six days can be expected to produce up to 50% less DNA compared to fresh sample processing *(26)*. A 30–40% decrease in DNA yield was observed in samples stored for three to seven days at 4°C, compared to fresh extractions *(26–29)*. Gustineich et al. *(27)* reported that DNA yield decreased by 30–40% if the blood was frozen at -20°C while Cushwa et al. *(29)* observed a 41% decrease in DNA yield in samples stored at -20°C. DNA yields of samples stored at -70°C were shown to be comparable to yields from samples immediately isolated *(26)*, *(30)*, *(31)*; however, Ross and coworkers observed a 25% decrease in DNA yield after similar storage of samples at -70°C *(32)*. In general, 150–250 µg of DNA can be isolated from 10 mL of whole blood *(33)*, although some laboratories have noted higher yields (e.g., 100–200 µg/5mL whole blood) *(34)*.

RNA can also be isolated from blood samples. For blood collection and preservation, PAXgene™ (PreAnalytiX GmbH, Switzerland) and CPT™ tubes (Becton Dickinson, NJ) have been widely used for whole blood and peripheral blood mononuclear cell (PBMC) collections *(35)* intended for RNA isolation. Extensive mRNA changes are eliminated or markedly reduced when whole blood is stored in preservatives contained in the PAXgene™ tubes *(36)*. The PAXgene™ system offers a number of potential advantages that makes it highly attractive for multicenter clinical studies, the primary one being ease of use *(36)*. However, some investigators have shown *(37)* that there is increased noise and reduced responsiveness in the gene expression profiles derived from whole blood compared with a leukocyte isolation protocol. These authors concluded that erythrocytes or reticulocytes and other nonleukocyte sources contribute an appreciable number of mRNA species in the whole blood collection system; but that simply removing the overabundant hemoglobin mRNA species does not result in a response pattern identical to that seen from leukocyte isolations.

The goal of any RNA isolation procedure is to recover an RNA population that mirrors the biology of the sample at the time of collection. Problems associated with the extraction of biologically representative RNA arise primarily from the susceptibility of RNA to degradation by ubiquitous and catalytically potent RNases. Therefore, RNA preservatives should be added, since many RNA transcripts change gene expression levels when stored (e.g., in EDTA) within hours (36, 38–40). It is important to note that the purity of the RNA as measured by A260/A280 is very consistent, even after extended storage of whole blood at ambient temperatures; and often the intactness of ribosomal RNA bands is also well maintained, although the underlying representation of many genes may have changed dramatically *(41)*. As noted previously, it may be important to reduce the globin mRNA population *(42)* contributed from the reticulocytes portion of whole blood samples, especially since the globin mRNA can contribute significantly to background noise in microarray experiments.

4.2 Tumor Biopsies

Paraffin embedded tumor samples are also utilized for pharmacogenomic studies. Formalin fixation and paraffin embedding is the standard tissue processing method used in many histopathology laboratories. This method allows for permanent preservation of tissues, easy storage, and optimal histological quality. However, formalin fixation may compromise the analysis of biomolecules, including DNA, mRNA, and proteins. Fresh frozen or immediately preserved tumor samples are preferred; however, samples prepared in this manner are not widely available.

A problem with the analysis of tumor samples is contamination of the samples by stromal cells (e.g., fibroblasts, myofibroblasts, and endothelial cells) and the surrounding normal cells. Even the most sophisticated genetic testing methods will be of limited value if the input material (nucleic acids) is not derived from sufficiently pure populations of the cells of interest. To address this problem, Emmert-Buck and colleagues introduced the laser capture microdissection (LCM) system *(43)* in 1996. LCM can be used to specifically obtain tumor tissue from surrounding normal tissue, whereby each laser pulse selectively transfers one small focal region of tissue or cell cluster to film contained on a slide *(43)*. This methodology does not adversely affect the ability to perform polymerase chain reaction (PCR) or other enzyme activity assays *(43)*. The success of LCM is illustrated by the large number of studies utilizing this technique for a broad range of downstream applications, such as loss of heterozygosity analysis (LOH), comparative genome hybridization (CGH) array analysis, methylation specific PCR, real-time (RT) quantitative (q)-PCR, expression microarrays, cDNA library construction, etc. In oncology, the genetic analysis of premalignant lesions has potential clinical implications, since these mutations represent an intermediate step of tumor progression from normal cells to cancer and may provide information with respect to malignant transformation. Analysis of these samples may also allow identification of multiple mutations (signatures or classifiers) that are associated with response to drug treatment.

RNA analyses can also be performed on samples obtained from LCM sections; however, the RNA yield is generally low and control of RNA quality is necessary to avoid misinterpretation of the gene expression results. Additionally, the elevated temperatures required for paraffin embedding are known to reduce the quality and yields of RNA. The use of different fixatives also has a significant effect on RNA integrity. Ethanol fixation and paraffin embedding of tissue specimens is not optimal for high-throughput mRNA expression analysis *(44)*; however, RT-PCR for specific genes can be performed on these samples. Kim et al. *(45)* showed that methacarn, a combination of methanol, chloroform, and acetic acid, was the optimal fixative for RNA studies; while Vincek et al. *(46)* showed that RNA can be adequately preserved in a new universal molecular fixative (UMFIX, Sakura Finetek USA, Inc., Torrance, CA). Other factors that can alter the integrity of RNA are the age of the paraffin block and the length of time that the samples have been stored. RNA extracted from archived FFPE blocks that are older than 10 years is typically about 100 nucleotides in length. However, newer microarray designs for genome-wide profiling of FFPE

samples from vendors such as Affymetrix allow the interrogation of smaller target sequences compared to standard gene chip arrays *(42)*.

Fresh tumor tissue can also be preserved by flash freezing in liquid nitrogen or by the use of RNALater™ (PreAnalytix GmbH, Switzerland). RNALater is more convenient for multicenter clinical trials, since tissue can be stored in RNALater at room temperature for up to three days without introducing any systematic changes in gene expression as measured in microarray experiments *(47)*. Protection of RNA in tumor samples has been previously accomplished by immediate lysis using high concentrations of detergents and/or chaotropic agents and organic solvents such as TRI reagent® (Applied Biosystems, CA). These methods are complex to use at the point of care and suffer from low sample throughput. Flash freezing of samples in liquid nitrogen and transport on dry ice are impractical in most clinical settings as well.

4.3 Serum/Plasma (Proteomic Analyses)

Proteomic analyses often complement genomic analyses and include interrogation of the entire proteome or portions of the proteome. The impact of preanalytical variables, ranging from patient posture to sample timing and tube type, on the quality of laboratory results for many protein measurements is well recognized *(48)*. In addition, other preanalytical aspects, such as centrifugation (speed, time, and temperature), storage time and temperature, and exposure to freeze-thaw cycles, are important. The direct effect of tube additives such as silicones, surfactants, and plasticizers on some analyses may be factors as well *(49)*.

These issues were focused on in the HUPO Plasma Proteome Project *(50)*. Within this project, comparison of serum and plasma specimens was done with respect to the human proteome. Serum samples were clotted by glass/silica-based activation and plasma specimens were derived using the three most common anticoagulants, namely, potassium-EDTA, lithium-heparin, and sodium citrate. The effects of storage were tested under various time and temperature conditions, and it was found that no major differences were observed between storage at –20°C, –80°C, and liquid nitrogen over two months time as detected by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF); however, there were differences at both room temperature and refrigerated storage. Since serum generation relies on a biochemical process, it is reasonable to expect that various parameters, such as temperature after sample collection, time for sample processing/clot formation, or medication of patients, can alter the peptide content of serum. These issues are difficult to standardize in routine clinical practice.

Therefore, the use of serum samples for peptidomic mono/oligo-biomarker discovery should be avoided in most cases. Serum peptide patterns have been used for prediction of early stage cancers, and a debate about this approach is ongoing *(51–54)*. At this time it is not clear whether the proteomic patterns reflect directly disease related peptides, or peptides that are generated due to secondary effects

during *ex vivo* coagulation. The choice of sample type is dependent on downstream analyses. Each of the individual sample types, serum, EDTA-plasma, heparin-plasma, and citrate plasma, exhibit shortcomings. EDTA forms soluble complexes with metal ions and should not be utilized if the endpoint measurement involves assays requiring divalent cations such as Mg2+ or Ca2+. Heparin can interfere in affinity processes such as SELDI-TOF analysis since it competes for binding of molecules to charged surfaces. Citrate can bind calcium and may falsely lower immunoassay measurements of multiple analytes *(55–57)*. Protease inhibitors may protect plasma proteins as early as phlebotomy procedures, and protease inhibitor use seems likely to provide a more reproducible sample. However, some inhibitors have the potential to alter proteins, and thus consideration of the desired analytical outcome is important.

5 Global Regulatory Considerations for Sample Collections

5.1 Country-Related Regulatory Considerations

Consideration of all the parameters discussed in the previous section should lead to standard collection procedures for sample procurement that will allow more robust downstream analyses. Emphasis on the quality measures taken for sample procurement may also assist with obtaining approval from relevant regulatory bodies for the intended genomic or proteomic research. As addressed in previous sections, many issues may be encountered in the pharmacogenomic protocol approval process. Other questions that may potentially arise relate to the processes surrounding sample acquisition and the degree of validation of the assay that will be used to analyze the samples. Efforts are currently ongoing to harmonize regulations for genomic sample collections; however, it may be well into the future before harmonization does occur. Because country-specific, local, and regional regulations continually change, it is recommended that one acquire and review specific country-related regulations prior to implementing a pharmacogenomic study at a particular site. Some countries regulate sample importation or exportation to stimulate commerce or to control data generation from ethnically derived sample sets. Some countries require importation or exportation application procedures that can be lengthy, and delays in sample procurement should be anticipated in these cases. Additionally, limits can be placed on the type of research, location of sample storage, coding of samples, and the rights to sample data. The following barriers can be expected to be encountered, either because of local regulations superseding country-level regulations or because of differences in the interpretation of specified regulations:

- Limitations on exploratory research: Argentina, Canada
- Prohibition of mandatory research: Korea, Spain
- Requirements for decriptions of research: Chile (requires gene listing)
- Requirements for a separate protocol describing research: Brazil, Thailand

- Separate approval bodies that require extensive time for review and approval: Australia, France, Israel, Netherlands
- Prohibition of anonymization: Brazil, Italy
- Limitations to location for sample banking: Iceland, Sweden
- Limitations to length of storage time: Italy, Netherlands
- Prohibition of sample storage: Malaysia, Taiwan, Thailand
- Requests for length of sample storage: Australia, Belgium
- Limitations or applications necessary for export: China, India, Spain
- Limitations to length of time samples can be outside country: Sweden
- Allowances for subject to request results of research: Brazil

5.2 Ongoing Efforts for Education and Policy Change Related to Sample Acquisition

To address some of these specific issues concerning global sample acquisitions, various groups have emerged to provide information in public forums that may assist in leading to harmonization of regulations across countries. These groups include (but are not limited to) the Pharmacogenetics Working Group, the European Federation of Pharmaceutical Industries and Associations, the Pharmacogenetics Research Network, the Council for International Organizations in Medical Sciences Working Group on Pharmacogenetics, the AAPS Pharmacogenetics and Pharmacogenomics Focus Group, and the Pharmacogenetics for Every Nation Initiative. The activities of these groups are described below.

The Pharmacogenetics Working group (PWG, http://www.pharmacogenetics working group.org) is a voluntary and informal association of pharmaceutical companies engaged in research in the science of pharmacogenetics. This group initially formed in response to regulatory requests for noncompetitive information from the industry. It provides information intended to promote a better public understanding of pharmacogenetic research and its development *(19, 24, 58, 59)*. The PWG works with the U.S. FDA, the EMEA, and regulators and various policy groups to provide information on noncompetitive issues related to pharmacogenetic research. The European Federation of Pharmaceutical Industries and Associations (EFPIA) has a pharmacogenomics task force. There is overlap in the membership between the EFPIA task force and the PWG. This task force does not currently have a separate website.

The Pharmacogenetics Research Network (http://www.nigms.nih.gov/pharma cogenetics, http://www.pharmgkb.org/), associated with NIH-NIGMS (National Institutes of Health-National Institute of General Medical Science), enables a network of multidisciplinary research groups to conduct studies addressing research questions in pharmacogenetics and pharmacogenomics in order to ultimately populate a knowledge base (PharmGKB) with data. The long-term goal of this group is to translate this knowledge and identify safe and effective drug therapies for individual patients. Among its other goals is to interact with and influence the wider community of

scientists in academia, industry, and government regulatory agencies in order to advance the field of pharmacogenetics.

Another group of interest is the Council for International Organizations in Medical Sciences, Working Group on Pharmacogenetics (CIOMS, http://www.cioms.ch). Of note, CIOMS has issued the "International Ethical Guidelines for Biomedical Research Involving Human Subjects" (developed in conjunction with WHO), which was published in 1993. The Working Group on Pharmacogenetics, which includes senior scientists from ten drug regulatory authorities and ten pharmaceutical companies, plus experts from WHO and academia, formed to consider drug development and the regulatory, ethical, educational, and economic issues related to pharmacogenetics. The findings and recommendations of the CIOMS-WGP have been presented at many international conferences in Europe, Japan, and the U.S.

The goal of the AAPS Pharmacogenetics and Pharmacogenomics Focus Group (http://www.aapspharmaceutica.com/inside/focus_groups/PGX/index.asp) is to provide a forum for information exchange on developments in pharmacogenetics and pharmacogenomics. They do this by generating yearly themes in these areas in the *AAPS Journal*, and by organizing symposia, workshops, roundtables, and guest speaker programs. The goals are to develop a knowledge base in pharmacogenetics and pharmacogenomics research, and to facilitate communication between academia, biotechnology, genomics firms, pharmaceutical companies, and regulatory agencies.

Finally, the Pharmacogenetics for Every Nation Initiative (PGENI, http://pgeni.unc.edu) has formed with four goals: i) to enhance the understanding of pharmacogenetics in the developing world, ii) to help build local infrastructure for future pharmacogenetic research studies, iii) to provide guidelines for medical prioritization for individual countries using pharmacogenetic information, and iv) to promote the integration of genetic information into public health decision making processes.

These and many other organizations are working towards the goal of providing comprehensive knowledge in the fields of pharmacogenetics and pharmacogenomics. It is anticipated that the activities of these groups may influence regulations applicable to genomic research, leading to harmonization of those regulations for samples intended for genomic analysis, as described above. It is hoped that such harmonization will occur in the near future, especially considering that a vast amount of genomic information is being captured and interpreted with the intent of personalizing medicine, in order to reduce unnecessary adverse events and to increase drug efficacy in individual patients.

In conclusion, many regulatory and operational considerations should take precedence over study start-up activities, when the intent is to acquire as many samples of high quality as possible for pharmacogenomic analysis. As discussed above, specific issues that may be encountered in the pharmacogenomic protocol approval process include inquiries related to:

- the intended use of samples collected for pharmacogenomic analyses,
- the length of time samples will be stored,
- sample coding procedures,

- management of the data collected,
- the maintenance of subject privacy and confidentiality,
- the physical sample storage location and the conditions under which samples are stored,
- allowance for and limitations on withdrawal of consent and sample destruction,
- limits on access to the sample data,
- reporting of results to individual subjects (and potentially genetic counseling),
- publication policies and the dissemination of results.

Since country-specific, local, and regional regulations continually change, it is recommended that one acquire and review specific regulations prior to placing a pharmacogenomic study. In addition, emphasis should be placed on standardizing global sample acquisition and handling procedures to ensure acquisition of samples of the highest quality and integrity for all intended downstream genomic applications.

References

1. EMEA, http://www.emea.europa.eu/
2. FDA, http://www.fda.gov/cder/
3. MHLW, http://www.mhlw.go.jp/english/index.html
4. PMC, http://www.personalizedmedicinecoalition.org/communications/overview.php
5. Eriksson S, Helgesson G (2005) Potential harms, anonymization, and the right to withdraw consent to biobank research. Eur J Hum Genet 13(9):1071–6.
6. Rivera R, Borasky D, Rice R, Carayon F, Wong E (2007) Informed consent: an international researcher's perspective. Am J Public Health 97(1):25–30.
7. Miller CK, O'Donnell DC, Searight HR, Barbarash RA (1996) The Deaconess Informed Consent Comprehension Test: an assessment tool for clinical research subjects. Pharmacotherapy 16(5):872–8.
8. Daugherty C, Ratain MJ, Grochowski E, et al (1995) Perceptions of cancer patients and their physicians involved in phase I trials. J Clin Oncol 13(5):1062–72.
9. Flesch R (1978) A new readability yardstick. J Appl Psychol 32:221–33.
10. Flesch R (1979) How to write plain English. Harper, New York.
11. Baker MT, Taub HA (1983) Readability of informed consent forms for research in a Veterans Administration medical center. JAMA 250(19):2646–8.
12. Duffy T, Kabance P (1982) Testing a readability writing approach to text revision. J Educ Psychol 74:733–48.
13. Eaton ML, Holloway RL (1980) Patient comprehension of written drug information. Am J Hosp Pharm 37(2):240–3.
14. Irwin J (1991) Teaching reading comprehension process, 2nd ed. Prentice Hall, New Jersey.
15. Vacca R, Vacca J (1989) Content area reading, 3rd ed. Scott Foresman, Illinois.
16. Kirsch I, Jungeblut A, Jenkins L, Kolstad A (1993) Adult literacy in America., U.S. Department of Education, Washington.
17. Regulations USCoF (2007) United States Code of Federal Regulations. In: 21 CFR 50.25.
18. Sciences CfIOoM (2002) International ethical guidelines for biomedical research involving human subjects. Geneva, ISBN 92-9036-075-5.
19. Anderson C, Gomez-Mancilla B, Spear B et al (2002) Elements of informed consent for pharmacogenetic research: perspective of the Pharmacogenetics Working Group.Pharmacogenomics J 2(5):284–92.
20. Wertz DC (2002) Genetic discrimination—an overblown fear? Nat Rev Genet 3(7):496.

21. Shickle D (2006) The consent problem within DNA biobanks. Stud Hist Philos Biol Biomed Sci 37(3):503–19.
22. Hoeyer K, Olofsson BO, Mjorndal T, Lynoe N (2004) Informed consent and biobanks: a population-based study of attitudes towards tissue donation for genetic research. Scand J Public Health 32(3):224–9.
23. ICH Harmonized Tripartite Guideline E6(R1): Good Clinical Practice: Consolidated Guideline, 1996.
24. Renegar G, Webster C, Stuerzebecher S et al (2006) Returning genetic research results to individuals: points to consider. Bioethics 20(1):24–36.
25. Bhutta ZA (2004) Beyond informed consent. B World Health Organ 82(10):771–7.
26. Madisen L, Hoar D, Holroyd C, Crisp M, Hodes M (1987) DNA banking: the effects of storage of blood and isolated DNA on the integrity of DNA. Am J Med Genet 27:379–90.
27. Gustincich S, Manfioletti G, Del Sal G, Schneider C, Carninci P (1991) A fast method for high-quality genomic DNA extraction from whole human blood. Biotechniques 11:298–300.
28. McCullough J, Carter S, Quie P (1974) Effects of anticoagulants and storage on granulocyte function in bank blood. Blood 43:207–17.
29. Cushwa W, Medrano J (1993) Effects of blood storage time and temperature on DNA yield and quality. Biotechniques 14:204–7.
30. Polakova H, Kadasi L, Zelinkova M (1989) The yield and quality of DNA extracted from blood samples stored under various conditions. Bratisl Lek Listy 90:844–7.
31. Bomjen G, Raina A, Sulaiman I, Hasnain S, Dogra T (1996) Effect of storage of blood samples on DNA yield, quality and fingerprinting: a forensic approach. Ind J Exp Biol 34:384–6.
32. Ross K, Haites N, Kelly K (1990) Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. J Med Genet 27:569–70.
33. Ellsworth D, Manolio T (1999) The emerging importance of genetics in epidemiologic research. I. Basic concepts in human genetics and laboratory technology. AEP 9(1):1–16.
34. McIndoe R, Linhardt M, Hood L (1995) Single tube genomic DNA isolation from whole blood without pre-isolating white blood cells. BioTechniques 19(1):30–2.
35. Fan H, Hedge P (2005) The transcriptome in blood: challenges and solutions for robust expression profiling. Curr Mol Med 5:3–10.
36. Rainen L, Oelmueller U, Jurgensen S et al (2002) Stabilization of mRNA expression in whole blood samples. Clin Chem 48:1883–90.
37. Feezor R, Baker H, Mindrinos M et al (2004) Whole blood and leukocyte RNA isolation for gene expression analyses. Physiol Genomics 19:247–54.
38. Haskill S, Johnson C, Eierman D, Becker S, Warren K (1988) Adherence induces selective mRNA expression of monocyte mediators and proto-oncogenes. J Immunol 140:1690–4.
39. Pahl A, Brune K (2002) Gene expression changes in blood after phlebotomy: implications for gene expression profiling. Blood 100:1094–5.
40. Tanner M, Berk L, Felten D, Blidy A, Bit S, Ruff D (2002) Substantial changes in gene expression level due to the storage temperature and storage duration of human whole blood. Clin Lab Haematol 24:337–41.
41. Davidson T, Johnson C, Andruss B (2006) Analyzing micro-RNA expression using microarrays. Method Enzymol 411:1–14.
42. Madabushi L, Latham G, Andruss G (2006) RNA extraction for arrays. Method Enzymol 411:1–14.
43. Emmert-Buck M, Bonner R, Smith P et al (1996) Laser capture microdissection. Science 274:998–1001.
44. Perlmutter M, Best C, Gillespie J et al (2004) Comparison of snap freezing versus ethanol fixation for gene expression profiling of tissue specimens. J Mol Diagn 6(4):371–7.
45. Kim J, Hwang M, Shin H et al (2003) Differential expression analysis using paraffin-embedded tissues after laser microdissection. J Cell Biochem 90:998–1006.
46. Vincek V, Nassiri M, Nadji M, Morales A (2003) A tissue fixative that protects macromolecules (DNA, RNA, and protein) and histomorphology in clinical samples. Lab Invest 90:1427–35.

47. Mutter G, Zahrieh D, Liu C et al (2004) Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. BMC Genom 5:88.
48. Guder W, Narayanan S, Wisserh H, Zawta B (2003) Samples: from the patient to the laboratory: the impact of preanalytical variables on the quality of laboratory results, 3rd ed. Wiley, New York. p 106.
49. Drake S, Bowen R, Remaley A, Hortin G (2004) Potential interferences from blood collection tubes in mass spectrometric analysis of serum polypetides. Clin Chem 50:2398–401.
50. Rai A, Gelfand C, Haywood B et al (2005) HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5:3262–77.
51. Villanueva J, Philip J, Entenberg D et al (2004) Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. Anal Chem 76:1560–70.
52. Petricoin E, Ardekani F, Hitt A, Levine B (2002) Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359:572–7.
53. Check E (2004) Proteomics and cancer: running before we can walk. Nature 429:496–7.
54. Diamandis E (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool. Proteomics 3:367–78.
55. Rai A, Stemmer J, Zhang Z et al (2005) Analysis of human proteome organization plasma proteome project (HUPO PPP) reference specimens using surace enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectrometry: multi-institution correlation of spectra and identification of biomarkers. Proteomics 5:3467–74.
56. Haab B, Geierstanger B, Michailidis B et al (2005) Immunoassay and antibody microarray analysis of the HUPO plasma proteome project reference specimens: systematic variation between sample types and calibration of mass spectrometry data. Proteomics 5:3278–91.
57. Tubes and Additives for venous blood specimen collection: approved standard-fifth edition, NCCLS. In. Wayne, PA; 2003:33.
58. Spear B, Heath-Chiozzi M, Barnes D, Cheeseman K, Shaw P, Campbell D (2001) Terminology for sample collection in clinical genetic studies. Pharmacogenomics J 1:101–3.
59. Lesko L, Salerno R, Spear B et al (2003) Pharmacogenetics and pharmacogenomics in drug development and regulatory decision making: report of the first FDA-PWG-PhRMA-Drug Safe Workshop. J Clin Pharmacol 43(4):342–58.