Published in final edited form as:

Pharmacogenomics. 2013 January ; 14(1): 1–4. doi:10.2217/pgs.12.188.

The architecture of pharmacogenomic associations: structures with functional foundations or castles made of sand?

Dylan M. Glubb¹ and Federico Innocenti^{1,2}

¹Eshelman School of Pharmacy, Institute for Pharmacogenomics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

²Lineberger Comprehensive Cancer Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Keywords

GWAS; eQTL; ENCODE; functional genomics; reporter gene assay; allelic expression imbalance; isogenic models; cancer genome

In the construction of any building, a foundation beneath the surface is required to ensure that the structure has a stable footing and is serviceable. Just like a building, a pharmacogenomic association requires a foundation. In this case, a foundation can be constructed from the elucidation of the genetic mechanism behind the pharmacogenomic association, providing a supportive biological basis. If the function of the genetic variant believed to provide the association (which we will call ‘associated SNP’ for simplicity) is not known, it is difficult to make a biological interpretation and determine whether the causal variant has been identified [1]. Consequently, establishing a functional foundation is a crucial step in the validation of any pharmacogenomic association and even more so when the ability to conduct external replication is limited. This is particularly a problem of pharmacogenomic studies stemming from clinical trials of experimental drugs, as independent study cohorts will very likely not exist. In these instances, the establishment of a functional foundation is an essential part in the process of demonstrating clinical validity and utility.

For candidate gene studies when SNP functionality is not known and for genome-wide association studies (GWAS), lack of efforts to prove functionality of associated SNPs severely limits their translation into clinically actionable markers. Decades of such studies have resulted in very few markers of clinical utility [2], and we cannot afford to repeat this experience in the future. The reality is that the vast majority of pharmacogenomic studies do

Correspondence: Federico Innocenti, MD, PhD, University of North Carolina at Chapel Hill, Eshelman School of Pharmacy, Institute for Pharmacogenomics and Individualized Therapy, 120 Mason Farm Road, Chapel Hill, NC, 27599, USA, innocent@email.unc.edu.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

not even attempt to provide a functional foundation for associated SNPs. For example, out of 23 pharmacogenomic GWAS [3], only three studies described any analysis of the function of associated SNPs [4-6]. The only information gained from two of the studies was that the genotype of the associated SNP correlated with expression of a candidate gene [5,6], which is not enough information to provide a functional foundation. The remaining study showed that ximelagatran was able to inhibit binding of peptides to the variant MHC class II protein encoded by *HLA-DRB1*0701* but not to variant proteins encoded by other *HLA-DRB1* alleles [4]. This is precisely the sort of information required to construct a functional foundation for a pharmacogenomic association: evidence of a SNP altering the pharmacology of a drug in a model system which reflects the clinical phenotype.

In GWAS, we also face the difficulty of putting into a functional context associated SNPs which are intronic, intergenic, or located in gene deserts as nearly 90% of associated SNPs from GWAS have been reported to be either intergenic or intronic [7]. Strategies to construct a functional foundation of cancer risk GWAS loci have been proposed [8], but very recent developments in functional annotation of the genome and the peculiar features of pharmacogenomics require the development of a different framework.

Isolating the causal genetic variant underlying a pharmacogenomic association is not a trivial task. The first step in this process, for any associated SNP, is to determine whether there are other genetic variants in linkage disequilibrium (LD). Knowledge of regional LD can be obtained through fine mapping and/or imputation methods using sequencing data from the 1,000 Genomes Project. Assuming the associated SNP is common (candidate gene studies and GWAS are usually designed to genotype common SNPs), the contribution of rare variants to the phenotype of interest should be also evaluated. Determining the clinical utility of rare variants in pharmacogenomics is presently problematic as there is no consensus as to an approach to combine rare variants to increase power in association studies. Non-synonymous rare variants have been grouped by their predicted effects on protein function, but the methods used to classify such variants have been rather limited and do not account for synonymous or non-coding rare variants [9]. Thus, rare variants should be incorporated into the downstream analyses we describe here to determine functionality and enable better categorization of such variants for association testing. In the near future, the challenges of functionally validating rare variants will increase with the output from genome sequencing.

After an expanded list of SNPs has been created, it should be determined whether the selected SNPs are located in a genomic regulatory region. There are many bioinformatic resources for this purpose available [10], but they should be complemented by the use of comprehensive human genomic data relating to the regulatory function of DNA from the Encyclopedia of DNA Elements (ENCODE) project [11]. Indeed, various ENCODE chromatin structure, histone modification and transcription factor binding information has been integrated with expression quantitated trait loci (eQTL) data, bioinformatic prediction and manual annotation into RegulomeDB, a web-based tool which categorizes genetic variants according to the available data [12]. RegulomeDB will be particularly useful in assessing the functionality of intergenic or intronic variants and it has been shown that this resource can identify putatively causal SNPs from GWAS [12].

Maps of eQTLs have been provided in multiple tissues, and probably represent the most studied annotation for functional inference of SNPs in the genome. eQTLs are SNPs associated with the interindividual variance in mRNA expression, and are more likely to associate with phenotypes in GWAS than other SNPs with matching frequencies [13]. Using eQTL information will aid the interpretation of the directionality of the postulated SNP effect (i.e., through increased or reduced mRNA expression of a nearby or more distant gene, even located on a different chromosome [13]), and can also bring attention to novel candidate genes. eQTL analysis can clarify the function of intergenic SNPs, as it is difficult to determine which gene the SNP might exert its effect through, especially if the SNP is far from flanking genes and/or there is significant LD decay. Web-based resources such as the Genotype-Tissue Expression eQTL Browser provide searchable data repositories of eQTLs from lymphoblastoid cell lines (LCLs), brain and liver [14]. Although the cell or tissue type that is the most appropriate to use for interpretation may vary according to the drug and phenotype under investigation, the central role of the liver in drug elimination suggests that liver eQTLs are extremely relevant for pharmacogenomic studies [14,15]. A limitation of the eQTL approach is that the protein encoded by a gene, rather than the transcript, is the functional molecule and, although one may expect a correlation between the transcriptional and translational expression of genes, the transcriptome and proteome are not necessarily equivalent [16]. High-throughput approaches to accurately measure the entire human proteome are currently not well developed [14], although targeted groups of proteins can be quantified, allowing a limited set of protein QTLs to be identified[17].

The *in silico* evaluations described so far should guide the design of the most appropriate and suitable experiments to test functionality, where molecular and cellular phenotypes should be evaluated at baseline and after a drug challenge. To validate functionality at the gene level in cell or animal models, manipulation of gene expression (knock-down or over-expression) demonstrate how a candidate gene can affect a cellular phenotype [18]. However, these approaches validate genes, and not genetic variants located in these genes. At least in the cancer pharmacogenomics field, the study of genomically characterized LCLs has been rather popular and effects of SNPs on cytotoxicity of chemotherapy have been described [19,20]. Because LCLs are derived from B lymphocytes, they are not the ideal model for the study of the anti-tumor responses of chemotherapeutics. For this purpose, the Cancer Cell Line Encyclopedia has great utility as it contains drug exposure and genome-wide data for nearly 500 cancer cell lines, allowing SNPs that associate with cytotoxicity to be identified [21].

In SNP-based validation studies, the location and the nature of the SNP will dictate the experimental design. The effects of coding SNPs can be analyzed by transfecting cells with variant cDNA constructs and relevant cellular or molecular phenotypes assayed [22]. In the case of variants in regulatory regions, sequences of DNA encompassing these variants can be cloned into reporter gene plasmid constructs so the effect of allelic variants on reporter gene expression can be isolated [22]. To measure the effects of genetic variants on mRNA processing, splicing and turnover, allelic expression imbalance assays in heterozygous individuals can show unequal levels of transcript corresponding to allelic variants [23]. The mechanistic effect of a genetic variant on gene expression can be elucidated by a variety of

methods including electrophoretic mobility shift assays, protein binding microarrays and chromatin immunoprecipitation techniques [23,24]. Again, for pharmacogenomic SNPs, the context of a drug challenge is a requirement. Highlighting this point is the observation that expression differences between alleles in reporter gene assays are sometimes only observed after cells have undergone a treatment [25].

In this framework, there are key limitations to approaches which use ectopic gene expression. For reporter gene assays, the constructs containing associated SNP(s) are not in their natural genomic context or state. For cDNA studies, the level of gene expression may be far greater than that of the endogenous gene. The use of isogenic model systems provides solutions to these issues. Isogenic cells or animals differing solely at one genetic locus can be created using homologous recombination through zinc finger nucleases [26]. This approach enables the effects of a genetic variant to be examined *in situ* in their natural genomic setting, with the added benefit that gene expression levels are at their endogenous levels. A major drawback of eQTL and similar approaches is that LD allows only correlative studies. Isogenic models enable the effects of specific variants to be directly interrogated and provide extremely flexible systems in which differences between alleles can be measured for any appropriate cellular or molecular phenotype.

With all cellular models, there is the issue of the suitability of the tissue for study. Pharmacogenetic variants can have effects on gene expression which are tissue dependent. For example, a *VKORC1* SNP that associates with warfarin dose requirement was found to have no effect on gene expression in human heart or LCLs, but a significant effect was observed in liver tissue [23]. Thus, the appropriate tissue for analysis must be carefully selected after consideration of the pivotal biological and pharmacological pathways involved in a pharmacogenetic phenotype.

While results from GWAS by and large do not inform treatment decisions, results from screens of the cancer genome do. In clinical programs using sequencing of the cancer genome, candidate genes and the related targeted therapy are selected on the basis of the myriad of somatic alterations in a given tumor [20]. It is imperative that functional knowledge of such alterations makes this process more scientifically rational, minimizing the empiricism of assigning therapy without functional annotations of genes and genetic alterations. For this application, standardized, high-throughput functional assays (not yet available) should be able to assay hundreds or thousands of genomic candidates for oncogenic or tumor suppressor activities [27].

It is very difficult to say what size effect of genetic variants in experimental systems translates into a clinically relevant response. Indeed, it has been shown that larger effects in reporter gene assays do not correlate with greater odds ratios in corresponding clinical genetic associations, although the variants that associate with clinical phenotypes typically show consistent effects in *in vitro* assays [25]. Moreover, evidence from any single model system should not be considered in isolation and is generally not sufficient to provide a functional foundation [23]. Rather concordance of the effect of a genetic variant from complementary sources that allows the development and testing of rational hypotheses should be considered of far greater importance in building a solid functional foundation.

Otherwise, pharmacogenomic associations will be like the castles made of sand that Jimi Hendrix sang about and “... fall in the sea, eventually”.

References

1. Sadee W. Pharmacogenomic biomarkers: validation needed for both the molecular genetic mechanism and clinical effect. *Pharmacogenomics*. 2011; 12(5):675–680. [PubMed: 21619429]
2. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med*. 2011; 364(12):1144–1153. [PubMed: 21428770]
3. Daly AK. Genome-wide association studies in pharmacogenomics. *Nat Rev Genet*. 2010; 11(4):241–246. [PubMed: 20300088]
4. Kindmark A, Jawaaid A, Harbron CG, Barratt BJ, Bengtsson OF, Andersson TB, Carlsson S, Cederbrant KE, Gibson NJ, Armstrong M, Lagerstrom-Fermer ME, Dellsen A, Brown EM, Thornton M, Dukes C, Jenkins SC, Firth MA, Harrod GO, Pinel TH, Billing-Clason SM, Cardon LR, March RE. Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J*. 2008; 8(3):186–195. [PubMed: 17505501]
5. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet*. 2009; 41(10):1105–1109. [PubMed: 19749757]
6. Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate M, Bassendine M, Spengler U, Dore G, Powell E, Riordan S, Sheridan D, Smedile A, Fragomeli V, Müller T, Bahlo M, Stewart G, Booth D, George J. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet*. 2009; 41(10):1100–1104. [PubMed: 19749758]
7. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106(23):9362–9367. [PubMed: 19474294]
8. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, James M, Liu P, Tichelaar JW, Vikis HG, You M, Mills IG. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*. 2011; 43(6):513–518. [PubMed: 21614091]
9. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet*. 2010; 86(6):832–838. [PubMed: 20471002]
10. Glubb DM, Paugh SW, van Schaik RHN, Innocenti F. A guide to the current web-based resources in pharmacogenomics. *Methods Mol Biol*. 2012 In Press.
11. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489(7414):57–74. [PubMed: 22955616]
12. Boyle A, Hong E, Hariharan M, Cheng Y, Schaub M, Kasowski M, Karczewski K, Park J, Hitz B, Weng S, Cherry J, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012; 22(9):1790–1797. [PubMed: 22955989]
13. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010; 6(4):e1000888. [PubMed: 20369019]
14. Glubb D, Dholakia N, Innocenti F. Liver expression quantitative trait loci: a foundation for pharmacogenomic research. *Front Genet*. 2012; 3:153. [PubMed: 22912647]
15. Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez J, Liu W, Lin YS, Moloney C, Aldred SF, Trinklein ND, Schuetz E, Nickerson DA, Thummel KE, Rieder MJ, Rettie AE, Ratain MJ, Cox NJ, Brown CD. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet*. 2011; 7(5):e1002078. [PubMed: 21637794]

16. Nie L, Wu G, Culley DE, Scholten JC, Zhang W. Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Crit Rev Biotechnol*. 2007; 27(2):63–75. [PubMed: 17578703]
17. Cyr DD, Lucas JE, Thompson JW, Patel K, Clark PJ, Thompson A, Tillmann HL, McHutchison JG, Moseley MA, McCarthy JJ. Characterization of serum proteins associated with IL28B genotype among patients with chronic hepatitis C. *PLoS One*. 2011; 6(7):e21854. [PubMed: 21750736]
18. Mullenders J, Bernards R. Loss-of-function genetic screens as a tool to improve the diagnosis and treatment of cancer. *Oncogene*. 2009; 28(50):4409–4420. [PubMed: 19767776]
19. Watters JW, Kraja A, Meucci MA, Province MA, McLeod HL. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proc Natl Acad Sci U S A*. 2004; 101(32):11809–11814. [PubMed: 15282376]
20. Innocenti F, Cox NJ, Dolan ME. The use of genomic information to optimize cancer chemotherapy. *Semin Oncol*. 2011; 38(2):186–195. [PubMed: 21421109]
21. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jane-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P Jr, de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palascandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R, Garraway LA. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012; 483(7391):603–607. [PubMed: 22460905]
22. Glubb DM, Cerri E, Giese A, Zhang W, Mirza O, Thompson EE, Chen P, Das S, Jassem J, Rzyman W, Lingen MW, Salgia R, Hirsch FR, Dziadziuszko R, Ballmer-Hofer K, Innocenti F. Novel functional germline variants in the VEGF receptor 2 gene and their effect on gene expression and microvessel density in lung cancer. *Clin Cancer Res*. 2011; 17(16):5257–5267. [PubMed: 21712447]
23. Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadee W. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood*. 2008; 112(4):1013–1021. [PubMed: 18523153]
24. Glubb DM, Innocenti F. Mechanisms of genetic regulation in gene expression: examples from drug metabolizing enzymes and transporters. *Wiley Interdiscip Rev Syst Biol Med*. 2011; 3(3):299–313. [PubMed: 20865777]
25. Ioannidis JP, Kavvoura FK. Concordance of functional in vitro data and epidemiological associations in complex disease genetics. *Genet Med*. 2006; 8(9):583–593. [PubMed: 16980815]
26. Di Nicolantonio F, Arena S, Gallicchio M, Bardelli A. Isogenic mutant human cells: a new tool for personalized cancer medicine. *Cell Cycle*. 2010; 9(1):20–21. [PubMed: 20016269]
27. Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. *Nat Med*. 2011; 17(3):297–303. [PubMed: 21383744]