Accelerating Drug Development with 23andMe Phenome-Wide Association Studies

J Fah Sathirapongsasuti, Brian T Naughton, Joanna L Mountain, David A Hinds, Joyce Y Tung, Cory Y McLean

23andMe, Inc., Mountain View, CA, USA

Introduction

Human genetics can be a powerful tool in the efforts to streamline drug development and prioritize research programs. When viewed through the lens of "experiments of nature," association between traits and naturally occurring gene mutations in humans can be used to predict the efficacy and toxicity of drugs targeting the gene products. This can yield valuable insights both through identification of likely successful drug targets and through reduced experimentation on drugs that ultimately fail to reach regulatory approval.

Two retrospective studies demonstrated the utility of human genetic data in drug development [1,2]. Using public [1] and private [2] genome-wide association study (GWAS) catalogs in conjunction with databases of approved or discontinued drugs, both studies highlighted examples of positive prediction of ultimately successful drug trials including Simvastatin, an LDL-lowering statin drug targeting the *HMGCR* gene. However, neither study made a quantitative statement about the predictive accuracy of this approach, possibly due to the paucity of positive predictions. Recently, protein-protein interaction network analysis has been used to further extend the scope and use of genetic association results in drug target validation [3]. A study of rheumatoid arthritis (RA) showed that interaction partners of genes associated with RA are 2.2 times more likely to be successful drug targets than the RA-associated genes themselves [4]. Thus, the combination of genetic associations and protein-protein interactions presents an appealing approach to target validation.

Methods

The applicability of the 23andMe GWAS catalog to this drug target validation approach has yet to be demonstrated. To assess the predictive accuracy of 23andMe database, we combined a public drug database, Therapeutic Target Database (TTD) [5], with 23andMe GWAS results for 308 curated medical phenotypes (September 2014). **Figure 1** depicts the overview of the method. TTD contains 5,951 drugs in various regulatory stages (1,985 approved, 24 withdrawn, 356 failed, 2,967 in trials, and 619 unknown). We used modified gestalt pattern matching [6] to link 23andMe phenotypes to drug indications via medical subject headings (MeSH). In total, 183 phenotypes were successfully linked to 852 indications through 456 MeSH terms (**Figure 1**). For the purpose of statistical assessment, only drugs with a terminal approval status (*approved, failed,* or *withdrawn*) were included and drugs targeting the same genes and phenotypes were grouped into drug classes. A total of 2,751 distinct drug classes had indications matching a 23andMe phenotype. Drug classes were considered "validated" if any SNP within the gene body was associated with the corresponding 23andMe phenotype above a specified significance threshold (**Table 1**). Additionally, as described in [4], genetic association validations were extended to interactors of the GWAS genes in the InWeb protein-protein interaction (PPI) network [7].

Acknowledgments

We thank 23andMe customers who consented to participate in research for enabling this study. We also thank employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

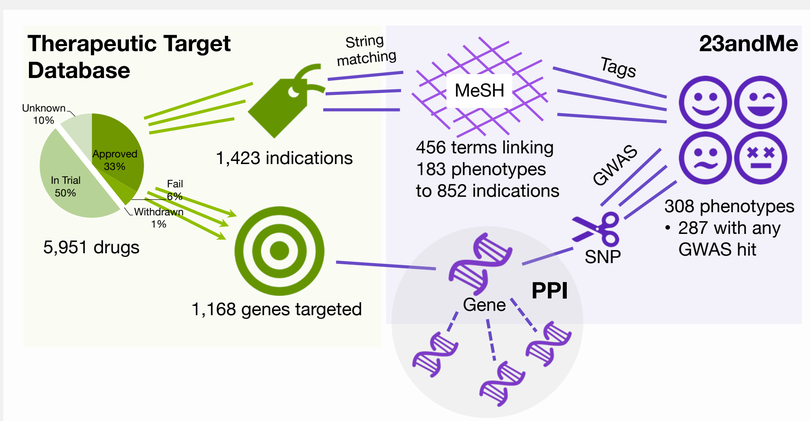

Figure 1. Overview of drug target validation by 23andMe GWAS catalog. Drug indications were linked to 23andMe phenotypes through medical subject headings (MeSH). Drug target validation is defined by significant association between targeted genes and matching phenotypes. Additionally, interactors of GWAS genes in a protein-protein interaction (PPI) were used to extend the genetic associations.

Table 1. Summary statistics of the target validation methods. GWAS significance levels $1x10^{-10}$, $5x10^{-8}$, and $1x10^{-6}$ all yielded strong enrichment of successful drugs. OR, odds ratio; PPI, protein-protein interactions; PPV, positive predictive value; NPV, negative predictive value.

* Examples given in Tables 2A (without PPI) and 2B (with PPI).

# Drug Classes	Approve	Fail	Total	Approve/ Fail OR	Fisher's p-value	PPV	NPV
With mappable indication	2070	681	2751	-	-	-	-
Validated at 1e-10 *	9	0	9	Inf	0.07699	1.000	0.248
Validated at 1e-10 PPI *	63	13	76	1.61	0.07219	0.829	0.250
Validated at 5e-8	13	0	13	Inf	0.02456	1.000	0.249
Validated at 5e-8 PPI	92	19	111	1.62	0.03309	0.829	0.251
Validated at 1e-6	16	0	16	Inf	0.01041	1.000	0.249
Validated at 1e-6 PPI	118	33	151	1.19	0.22780	0.781	0.249

Results

We observed strong enrichment of successful drugs among those validated by 23andMe GWAS (**Table 1**). At the conservative significance level $1x10^{-10}$, we validated 9 drug classes including statin drugs whose target, *HMGCR*, is significantly associated with high cholesterol phenotype ($p=9.2x10^{-37}$, **Table 2A**). Each of the 9 validated drug classes contain successful drugs. We observed the same enrichment even when the GWAS significance level was relaxed to $5x10^{-8}$ (13 drug classes validated; all successful) and $1x10^{-6}$ (16 drug classes validated; all successful). Moreover, when the PPI network was used to extend the associations, we observed a consistent trend toward enrichment with substantially more drug classes validated (examples shown in **Table 2B**). Some successful drug targets interact with GWAS genes but are not themselves associated with the condition. Concrete examples of this are *F10*-inhibitors, such as Enoxaparin (Sanofi) and Fondaparinux sodium (GSK), used to prevent blood clots (venous thromboembolism). Factor X (encoded by *F10*) binds to Factors II and V (*F2* and *F5* genes) which are associated with "blood clots" phenotype in 23andMe GWAS ($p=1.3x10^{-24}$ and $p=3.6x10^{-137}$, respectively). Utilizing the PPI network, we were able to validate *F10*-inhibitors through the association of *F2* and *F5* (**Table 2B**). Thus, in our retrospective study, drug target validation by 23andMe GWAS, with or without PPI, strongly correlates with successful drug development.

Discussion

Because of the rarity of genetic associations, even in a large GWAS catalog such as 23andMe's, the negative predictive value of our predictions remains low (NPV=0.248-0.251; **Table 1**). Consequently, the absence of a genetic association between a drug indication and a particular gene does not necessarily predict failure of a drug that targets that gene. The absence of a link between a 23andMe genetic association and a drug target could arise for multiple reasons, including 1) a poor matching of phenotype and drug indication, 2) drug mechanism of action not sufficiently captured by the genetic model (e.g. due to population structure) or 3) a lack of statistical power in the GWAS. On the other hand, linkage disequilibrium (LD) may induce spurious signals by linking GWAS associations to non-causal genes. Also, due to selection bias toward successful drugs (75.2%) in the drug database TTD, our positive predictive value of 78-100%, while representing a significant improvement over the baseline, should not be interpreted as the post-test probability of success for a drug targeting a gene with a significant phenotype association. Nonetheless, our results suggest benefit of using 23andMe GWAS catalog to guide drug target validation.

Table 2. (A) Examples of drugs validated by 23andMe GWAS. *HMGCR* is a well-established hypercholesterolemia risk genes; the finding indicates the validity of our approach. **(B) Examples of drugs validated by extending GWAS results through PPI network.** *F10*-inhibitors (e.g. Enoxaparin) are validated through associations of *F2* and *F5* with the "blood clots" phenotype.

In LD with multiple genes in the HLA locus.

A. Without PPI

Target	23andMe Phenotype	Indication	Drug	Company	Status	GWAS p-value
HMGCR	**High Cholesterol**	Hypercholesterolemia	Atorvastatin	Pfizer	Approved	9.20e-37
F2	**Blood Clots**	Venous Thromboembolism	Ximelegatran	AstraZeneca	Approved	1.33e-24
HBB	**Anemia**	Iron Deficiency	Iron Dextran	None	Approved	3.65e-19
TNF	**Autoimmune**	Rheumatoid Arthritis	Etanercept	Amgen	Approved	7.93e-18
ADRB1	**High Blood Pressure**	Hypertension	Alprenolol	AstraZeneca	Approved	4.50e-17
ADRB1	**CVD**	Hypertension	Alprenolol	AstraZeneca	Approved	9.74e-17
HBB	**Blood Disorder**	Iron Deficiency	Iron Dextran	None	Approved	1.98e-14
TPO	**Hypothyroidism**	Hyperthyroidism	Propylthiouracil	None	Approved	4.43e-11
GHR	**Height**	Growth Hormone Deficiency In Children	Somatropin Recombinant	Pfizer	Approved	7.52e-11

B. With PPI

Target	23andMe phenotype	Indication	Drug	Company	Status	GWAS p-value	GWAS gene
F10	**Blood Clots**	Venous Thromboembolism	Enoxaparin	Sanofi-Aventis	Approved	3.60e-137	F5
F2	**Blood Clots**	Venous Thromboembolism	Ximelegatran	AstraZeneca	Approved	3.60e-137	F5
HLA-DRB1	**Multiple Sclerosis**	Multiple Sclerosis	Glatiramer Acetate	Teva Pharma	Approved	2.37e-82	HLA-DRA#
MTTP	**High Cholesterol**	Hypertriglyceridemia	Implitapide	MRL Int'l	Failed in Phase II	2.56e-49	APOB
HMGCR	**High Cholesterol**	Hypercholesterolemia	Atorvastatin	Pfizer	Approved	9.20e-37	HMGCR
GHR	**Height**	Growth Hormone Deficiency In Children	Somatropin Recombinant	Pfizer	Approved	3.91e-34	SOCS2
ESR1	**Multiple Sclerosis**	Multiple Sclerosis	Estriol	None	Approved	2.45e-33	CSNK2B#
F2	**Blood Disorder**	Thrombocytopenia	Lepirudin	Bayer	Approved	2.03e-30	F5
SERPINC1	**Blood Clots**	Deep Vein Thrombosis	Tinzaparin	Leo Pharma	Approved	1.82e-27	F11

References

1. Kamb A, Harper S, Stefansson K. Human genetics as a foundation for innovative drug development. *Nat Biotechnol*. 2013 Nov;31(11):975-8.
2. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, Vincent M. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol*. 2012 Apr 10;30(4):317-20.
3. Zhao S, Iyengar R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu Rev Pharmacol Toxicol*. 2012;52:505-21.
4. Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014 Feb 20;506(7488):376-81.
5. Zhu F, Shi Z, Qin C, Tao L, Liu X, Xu F, Zhang L, Song Y, Liu XH, Zhang JX, Han BC, Zhang P, Chen YZ. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res*. 40(D1): D1128-1136, 2012.
6. Ratcliff JW, Metzener D. Pattern Matching: The Gestalt Approach, *Dr. Dobb's Journal*. 1988 July:46.
7. Lage K, Karlberg EO, Størling ZM, Olason PI, Pedersen AG, Rigina O, Hinsby AM, Tümer Z, Pociot F, Tommerup N, Moreau Y, Brunak S. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol*. 2007 Mar;25(3):309-16.