It's Time for a Human Knockout Project to Advance Drug Discovery

Human beings just might be the ultimate model organism, the ultimate drug discovery tool.

Read one way, lacking in context, a comment like that might sound glib or even dangerous. But bear with me. I'm really talking about understanding nature's experiment - genetic mutations that have occurred naturally over the course of human history. These germline genetic tools are increasingly becoming available, and they have the potential to turn humans into the ideal model organism for drug discovery research. It's time to think about a "Human Knockout Project," akin to the Human Genome Project, which I described last week in a Nature News & Views based on an original research article in *Nature* (see here, BBC *Inside Science* podcast here) and which has been championed by thought leaders such Daniel MacArthur of the Broad Institute and Massachusetts General Hospital (see 2014 *Science* story here, his 2012 blog here).

But before diving in, a bit about me. This is my first article for Timmerman Report, and I'd like to introduce myself (see my biography here) to provide context for readers of the Timmerman Report. I am trained as a physician-scientist. I practiced clinical rheumatology and led an academic genetics & genomics lab at Brigham & Women's Hospital, Harvard Medical School and the Broad Institute for more than 10 years before moving into industry. I am currently transitioning between a job at Merck (Head of Translational Medicine) to a new job in the biopharma industry based in Cambridge. I frequently blog (www.plengegen.com) and tweet (@rplenge) about human biology & drug discovery.

Now, some background on human knockouts and the application to drug discovery & development. At some point in the near future - I estimate 5 years - publicly available databases will contain detailed information on the phenotypic impact of genetic perturbations on the vast majority of genes in the human genome. Such genetic perturbations may come in the form of a subtle effect on gene function. One example, might be a two-fold change in gene expression for a common, non-coding variant identified via a genome-wide association study (GWAS). Or, the database might show a more extreme effect on gene function, such as complete gene ablation for a rare variant identified from sequencing a family with a Mendelian disorder. Assimilating all available data will provide nature's estimate of function-phenotype maps, and therefore mimic the effect required for pharmacological manipulation of the same target (see 2013 review in Nature Reviews Drug Discovery).

But these databases need to be built, as the existing databases are good but incomplete. GWAS catalogues are available. Open Targets and the NHGRI-EBI Catalog are a couple examples. The problem is that most public GWAS databases are not integrated in a manner required for drug discovery & development. There are also databases of gene mutations implicated in rare Mendelian diseases and human gene knockouts in individuals from the general population who have not been ascertained for a specific disease phenotype. These databases also have potential, but suffer from similar weaknesses that make them only marginally useful for drug discovery.

Just last week, Sekar Kathiresan, from the Massachusetts General Hospital, Harvard Medical School and the Broad Institute, together with colleagues from across the world, published an original research article in *Nature* that provides one of the largest available catalogues of

individuals from the general population who inherit gene knockouts. This publication nicely complemented other studies from Iceland (see Salem et al Nature Genetics 2015), the Exome Aggregation Consortium (ExAC, see Lek et al Nature 2016), and East London Genes & Health (see Narasimhan et al Science 2016). We are beginning to see there's a lot to learn from these "human knockouts." That is, these individuals were not studied because they manifest a common phenotype for a common disease like diabetes or rheumatoid arthritis. They weren't studied to specifically look for a rare disease phenotype such as cystic fibrosis or primary immune deficiency. By looking at human knockouts, it is possible to study the effect of a complete gene ablation on individuals who otherwise might be healthy, or who may currently appear healthy but face a biochemically increased risk of developing disease.

How might a fully integrated, aspirational genetic database - and human knock-out data in particular - be used for drug discovery & development? I see two broad applications: to *rule-in* and to *rule-out* drug targets. For human knock-out data, the key concept is to understand the effect of maximum genetic perturbation on human physiology.

(1) **Rule-in drug targets:** As has been described by Matt Nelson and colleagues from GlaxoSmithKline (see 2015 Nature Genetics), and David Cook and colleagues from AstraZeneca (see 2014 Nature Reviews Drug Discovery), therapeutic molecules developed against targets with human genetic data are more likely to lead to regulatory approval than those without. PCSK9 represents the poster child for human genetic knockouts in drug discovery & development (see my plengegen.com blog here). But there are many other examples, too. A particularly unusual story is that of 5-alpha reductase deficiency, which causes pseudo-hermaphrodism at birth, and small prostates and lack of male pattern baldness in adults. These observations led to the development of a 5-alpha reductase inhibitor, known as finasteride.

(2) **Rule-out drug targets**: But human genetics can also rule-out drug targets or mechanisms that are nominated through animal models, human epidemiology or other approaches. A prominent example is related to raising HDL cholesterol, the so-called "good cholesterol". For decades, human epidemiology has suggested that therapeutically raising HDL through pharmacological perturbations of targets such as cholesterylester transfer protein (CETP), will protect from cardiovascular disease. Human genetics, however, now clearly demonstrates that genetic variants that solely raise HDL do not protect people from cardiovascular disease. Multiple failed clinical trials of CETP-inhibitors support human genetics over human epidemiology. Similarly, the *Nature* study by Sekar Kathiresan and colleagues provide support that pharmacologically inhibiting the enzyme Lp-PLA2, which is an inflammatory enzyme expressed in atherosclerotic plaques, will *not* have cardioprotective effects. The *Nature* study identifies human knockouts of the gene, *PLA2G7*, that codes for Lp-PLA2 and demonstrates no protection from cardiovascular disease. This observation, which conflicts with the putative beneficial effects predicted from human epidemiological studies (see Lancet 2010), has also been born out in clinical trials, where an Lp-PLA2 inhibitor (darapladib) has not had a beneficial impact on cardiovascular disease (see NEJM 2014, JAMA 2014). Assuming future clinical trials with Lp-PLA2 inhibitors continue to show no cardioprotective effect, then the Lp-PLA2/*PLA2G7* story provides a triumph of human genetics over correlative studies via human epidemiology.

Towards this end, a road map for a Human Knockout Project is now coming into focus. The goals are simple: to identify human knock-outs for each of the approximately 20,000 genes

in the human genome and to understand the physiological consequences of these gene knockouts in the ideal model organism, humans. A Human Knockout Project will require sequencing individuals from the general population who have not been selected for a disease phenotype (as in Sekar Kathiresan's *Nature* study), as well as sequencing individuals selected for a specific clinical phenotype. Ideally, people with common diseases, rare diseases, and embryos that died in utero would all be part of the Human Knockout Project sequencing initiative.

Data from the sequencing of humans could be quite powerful for drug discovery. For example, approximately 20% of genes in the human genome, when knocked-out, lead to embryonic lethality; another 30% of genes, when knocked out, lead to one of over 7,000 rare Mendelian diseases. Complete genetic ablation of *JAK3*, which cause primary immunodeficiency, is another good example of a human knockout that led to the development of a successful medicine. Tofacitinib, which inhibits the JAK3 protein, is approved for treatment of the inflammatory conditions rheumatoid arthritis and psoriasis. Other genes - and the number is not known, but is likely at least 50% of protein-coding genes - will not lead to embryonic lethality or overt disease. Lose one of those non-essential genes, and you can get along just fine. Maybe these mutations increase your risk of disease or lead to subtle biochemical alterations that cannot be seen with the naked eye. Or more interesting, some gene knockouts may even protect from human disease. Those looking for a sequel to the PCSK9 story are probably aware that knockouts of APOC3 lower triglyceride levels and protect from cardiovascular disease. Sekar Kathiresan and colleagues describe the beneficial consequences of complete lack of APOC3 in their *Nature* article.

So, what to do? Based upon what we know from studies such as those performed by Sekar Kathiresan and colleagues, *millions of individuals* will need to be sequenced to identify at least one person with a human genetic knockout for each of the 20,000 genes in the human genome. While this may seem daunting at first, initiatives are underway which demonstrate feasibility.

Here are concrete steps that should be taken to advance a Human Knockout Project for drug discovery & development.

(1) **Exome sequencing should be performed with individuals from diverse outbred and inbred populations, without regard for any specific disease phenotype.** In addition to Iceland, ExAC, PROMIS, and East London Genes & Health, GSK and Regeneron have committed significant resources to sequence as many as 500,000 individuals from the UK Biobank. The US Precision Medicine Initiative, now known as the "All of Us" research program, aspires to sequence at least 1 million people. I propose that a successful Human Knockout Project would generate sequence data on at least 200,000 individuals from consanguineous parents (predicted to identify knockouts for ~40% of genes in the human genomes) and >2 million individuals from ethnically diverse outbred populations (predicted to identify knockouts for ~30% of genes). Sequence data alone is insufficient, however. It is critical that individuals sequenced also have longitudinal clinical data and be consented for recall for additional studies.

(2) **Exome sequencing for all known Mendelian disorders.** There are an estimated 7,000 rare Mendelian disorders. Approximately half of these disorders have at least one gene identified as disease-causing, although multiple genes are often implicated in these rare

diseases. A successful Human Knockout Project would sequence the majority of these Mendelian disorders to identify the causative gene.

(3) **Exome sequencing for the majority of common disorders in at least 25,000 cases**. Population genetics predicts that at least 25,000 cases (and a larger number of controls) will be required to identify disease-associated rare variants for common diseases (see Zuk et al PNAS 2014). Although many of these rare variants will be partial loss- or gain-of-function, some disease-associated rare variants will also be complete genetic knockouts. To date, very few, if any, common diseases have reached this sample size. A successful Human Knockout Project would expand exome sequence data for select number of common diseases in at least 25,000 cases per disease to establish proof-of-concept to support, or refute, population genetic theory.

(4) **Exome sequencing in embryos that do not survive to term.** In theory, it should be possible to predict which genes, when mutated, lead to embryonic lethality by sequencing living individuals. It would be useful, however, to also understand directly the correlation between human knockouts and nature of the developmental abnormality (e.g., defects in heart, immune system, brain). A successful Human Knockout Project would exome sequence a modest number of embryos, say several thousand, that failed to survive to term as a pilot to define value of an expanded effort for drug discovery and development.

(5) **Integrate human knockout data with other human genetic data in a publicly available genetic database.** As mentioned above, human knockout data represents one valuable source of data to establish genotype-phenotype correlations. The vast majority genetic variants that are disease-associated will not be human knockouts, but rather will be variants that have partial loss- or gain-of-function. Moreover, most disease-associated genetic variants from GWAS reside outside of protein-coding sequences. Therefore, a successful Human Knockout Project must combine human knockout data with GWAS and other genomic data - including data from model organisms - to derive the greatest value for drug discovery and development. The informatics challenge of such a database should not be underestimated.

Of course, these are still early days for large-scale genome sequencing. But the future is now. There's time to organize the public databases in a way to make them most useful for academic researchers, and for industrial drug discovery. If you have further questions or comments on how a Human Knockout Project might be most useful, please leave a comment at the end of this article or send me a note at Robert.plenge@gmail.com.