Review

CellPress

Human knockout research: new horizons and opportunities

Fowzan S. Alkuraya[1,2]

[1] Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia
[2] Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

Although numerous approaches have been pursued to understand the function of human genes, Mendelian genetics has by far provided the most compelling and medically actionable dataset. Biallelic loss-of-function (LOF) mutations are observed in the majority of autosomal recessive Mendelian disorders, representing natural human knockouts and offering a unique opportunity to study the physiological and developmental context of these genes. The restriction of such context to 'disease' states is artificial, however, and the recent ability to survey entire human genomes for biallelic LOF mutations has revealed a surprising landscape of knockout events in 'healthy' individuals, sparking interest in their role in phenotypic diversity beyond disease causation. As I discuss in this review, the potentially wide implications of human knockout research warrant increased investment and multidisciplinary collaborations to overcome existing challenges and reap its benefits.

In search of a medical understanding of gene function

The quest to understand the function of the human genome was the primary motive for the Human Genome Project, but more than a decade after the 'complete' sequencing of the human genome our knowledge of the function of its individual components remains limited. Notwithstanding the debate surrounding what constitutes the 'functional genome' [1–4], even the classical functional units of the genome, that is protein-coding genes, are far from being fully understood as most of their encoded proteins have no established developmental/physiological function. With nearly 20 000 protein-coding genes in the human genome, the overwhelming majority of which encode more than one protein by virtue of alternative splicing [5], assigning a function to each of these genes is clearly a daunting task. The acute requirement for high-throughput approaches to tackle this challenge catalyzed the growth of many branches of functional omics, and although a wealth of data has been generated as a result, its demonstrated medical relevance is at its nascent stages [6].

An alternative approach, founded on the premise that medically relevant genes should result in an abnormal phenotype when their function is perturbed, has provided most of the medically relevant functional annotation of human genes [7,8]. This approach has seen its most resounding success when perturbation of a single gene is sufficient to produce a tractable phenotype, which is the basis of Mendelian genetics. With a mutation rate of 1.2×10^{-8} per nucleotide per generation (not including the mutation rate at the copy number level) [9], and >130 million births per year, the human population provides a vast resource to study Mendelian genes and the entire spectrum of their phenotypic effects. Combined with the recent advent of powerful genomic analytic tools, Mendelian genetics is proving to be a high-throughput, medically relevant 'functional annotator' of the human genome. This explains the resurgence of interest in this field after it was overshadowed transiently by the study of complex genetics where unprecedented investment has assigned functionality to a frustratingly small number of genes despite the numerous potential targets that have been identified [10–12].

A special scenario in Mendelian genetics that is particularly useful in this regard is when an individual gene harbors a LOF mutation in both alleles, rendering the gene completely inactive, equivalent to an experimental knockout in model organisms. This is pervasive in autosomal recessive disorders and has long been leveraged as a benchmark against which all other classes of mutations are compared for their detrimental effect. Although the notion that human knockouts represent the ultimate *in vivo* environment to investigate gene function has far reaching implications, the potential of these naturally occurring experiments has largely been confined to the context of 'disease' and thus not fully unlocked. In part, this may be due to the fact that most human traits are quantitative, and therefore it is not straightforward to connect genotype to phenotype. Another important reason is the failure to view 'disease' as a subjective annotation of a particular phenotypic state that is more appropriately viewed as a phenotypic data point along a spectrum that ranges from early embryonic lethality to lack of phenotypic consequence and the entire range between these two extremes (Figure 1) [13]. The current ability to identify knockout events in every individual genome using new sequencing tools provides an unprecedented opportunity to move freely between the genotype and phenotype thus bypassing historical roadblocks in defining the true spectrum of phenotypic consequences of gene loss in humans. This review discusses many aspects of this exciting line of research with special emphasis on its potentially far-reaching medical implications. Existing

Corresponding author: Alkuraya, F.S. (falkuraya@Kfshrc.edu.sa).
Keywords: loss of function; autozygome; evolution; druggable targets; adaptation.

0168-9525/
 http://dx.doi.org/10.1016/j.tig.2014.11.003

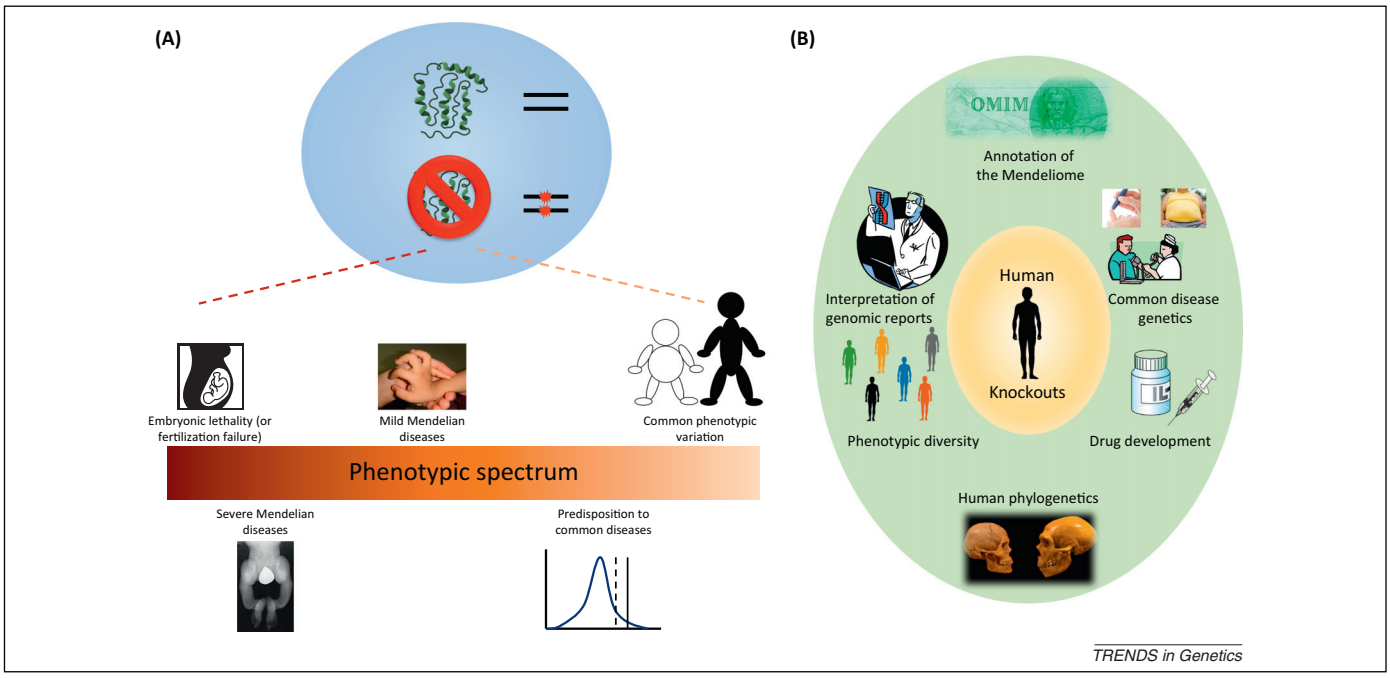

Figure 1. **(A)** Phenotypes associated with human knockout events can span the entire phenotypic spectrum and are not limited to diseases. **(B)** Research in human knockouts has a wide range of basic and translational applications.

and perceived challenges are also discussed with some solutions suggested.

The evolutionary legacy of LOF in the human genome

LOF mutations are defined as genetic (or genomic) alterations that render a particular allele completely inactive. Genomic rearrangements that result in whole gene deletion are conspicuous examples of LOF where the gene is physically removed [14]. More often, however, the null nature results from frameshift insertion or deletion or nonsense mutations giving rise to premature truncation of the encoded proteins. Canonical splice-site mutations usually result in alteration of the normal transcript, and although they do not necessarily result in a frameshift they are often counted among LOF mutations. Until recently, the magnitude of LOF mutations in the human genome was largely inferred from epidemiological studies that extrapolated their carrier frequency from analysis of autosomal recessive diseases (e.g., lethal equivalents) [15]. However, we now have empirical data based on whole-genome and whole-exome sequencing of a large number of individuals that revealed the presence of around 400 LOF variants per individual [16]. As I discuss later in detail, defining a variant as LOF is not straightforward so caution has to be exercised in interpreting this and related figures in the literature. Nonetheless, it is clear that LOF variants are remarkably more common than previously thought [17].

Most LOF variants are low in frequency, which is consistent with negative selection, but some exhibit a higher population frequency than would be expected for detrimental alleles, which raises interesting questions about the factors that maintain these variants at such high frequencies [18]. In its simplest form, LOF can reach fixation when it represents a human-specific 'pseudogenization' event. Differentiating passive from adaptive pseudogenization is paramount because the latter is particularly relevant to our understanding of how gene loss may have endowed humans with selective advantage, but doing so is not always easy. Inference of positive selection is usually based on such criteria as selective sweep with long haplotypes, short time from allele origin to its fixation, a dearth of polymorphisms in the null versus ancestral alleles, and pre- and post-admixture analysis of human populations [19–23], but these criteria are not foolproof. For example, local recombination hotspots can hide the long haplotype signature of positively selected LOF [24].

Adaptation through LOF is the basis of the 'less is more' hypothesis that suggests that loss of some genetic material can accelerate evolution [25]. This phenomenon has been experimentally observed in bacteria that adapt by dispensing of genes involved in the metabolism of nutrients no longer available to them, perhaps in an attempt to contain unnecessary cost associated with the expression of those genes and to rewire cellular networks to adapt to the new metabolic environment [26]. It has been suggested that LOF mutations may have contributed to the distinction between humans and other great apes although recent evidence is not consistent with excess of fixed LOF in the human branch [27]. Several attempts have been made to catalogue human lineage-specific pseudogenization events and these revealed interesting patterns of enrichment of certain classes of genes, including chemoreception and immune response genes, and a case has been made for positive selection of some of these genes, such as *CASPASE12*, deficiency of which has been shown to reduce the risk of sepsis [28–30].

More recent gene inactivation events in the human lineage can also reach a high population frequency due to several factors. Apart from bottlenecks and genetic drift, which can passively drive LOF alleles with deleterious health effects to a relatively high frequency as shown in