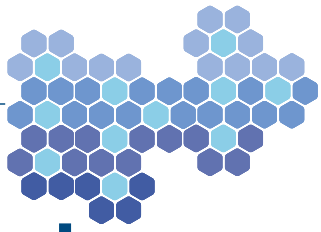

Going Forward

The next generation of gene therapy has a promising outlook. Improving therapeutic construct design with rationally-designed expression cassettes could revolutionise the field

By Dr Michael L Roberts at Synpromics

Gene therapy has come a long way since the early 2000s, when, after a period of intense excitement, it went through a bit of a hiatus. Essentially, a series of severe adverse events during several trials, including the death of a patient, forced the budding industry to reflect on how it was delivering therapeutic transgenes to patients. Over the past decade, researchers have developed safer gene delivery vectors that addressed some of these earlier issues, and several of these next generation vectors have been used in clinical trials with enormous success. This has resulted in the approval and near-approval of several new genetic-based medicines in the western hemisphere.

Thus, for many indications, the issues of gene delivery have been adequately addressed, and the bottleneck in the development of effective gene therapies lies now in how the expression of the therapeutic gene is regulated. This article notes how an enhanced understanding of the functional regulation of the human genome is enabling the development of improved rationally-designed gene therapeutic expression cassettes, thus helping the industry to build the future generation of cell and gene therapies.

Keywords

Expression cassettes
Gene promoters
Vectorology

The Next Generation

Cell and gene therapy has made a strong resurgence over the past few years, driven, in part, by some major successes in a diverse array of indications: ranging from the emergence of adoptive T-cell therapy as an effective treatment for certain lymphomas, the approval of GSK's *Strimvelis* for the treatment of ADA-SCID and the near approval of Spark Therapeutics' *Luxturna*, a gene therapy treatment for Leber congenital amaurosis. Investors are now flooding back to the field, and an enormous amount of capital has been injected into the sector as companies seek to exploit academics' recent advances.

Successful gene therapy of any disorder can only be achieved if three main hurdles are overcome: delivery of the therapeutic cassette to the target cells, sufficient expression of the desired gene to the required levels and restricted to the appropriate cells and the evasion of an immune response to the delivery vehicle. Significant effort has gone into the development of novel vector systems so that the target gene can be delivered to enough cells to elicit a therapeutic effect. Likewise, new promoters now need to be isolated that can drive efficient, yet finely tuned, gene expression, allowing the therapeutic effect to last for the individual's lifetime. Finally, given that the target gene is often delivered using immunogenic vectors (both viral and cell) and the therapeutic gene is often new to self, long-term immune responses to the therapeutic cassette need to be minimised.

Significant advances have been made in vectorology, where several groups have successfully adapted vector systems, such as lentiviral and adeno-associated viruses (AAV), for gene delivery. Capsid engineering of AAV vectors has resulted in the creation of a range of vectors that have differing tropisms and can safely deliver payloads to diverse cell types and tissues. Addressing the issues of expression and immunity can be achieved through the creation of novel promoters. Ideally, expression of therapeutic genes should be fine-tuned so that the protein is expressed to appropriate levels that yield a long-term therapeutic effect while simultaneously preventing expression in antigen-presenting cells to minimise host immune response.

Optimal Gene Expression

In the next phase of gene therapy development, it is essential that the impact of factors related to expression of the therapeutic transgene are mapped out. Specifically, how variations in the efficacy of gene expression impact which patients respond to the therapy and which do not. Gene therapy is often heralded as a curative modality, however, for this to be true, the gene must be expressed for a patient's lifetime. Therefore, it is imperative that the promoters driving that expression are not silenced.

The appropriate expression strength of the therapeutic gene

needs to be determined to avoid any toxicity associated with over-expression and, conversely, reduce the likelihood of sub-therapeutic expression dose. To avoid off-target effects and improve safety, the therapeutic gene should be expressed solely in the target cell or in the pathologic tissue. Finally, a built-in mechanism that enhances the dialling-up or -down of expression is required to ensure expression of the therapeutic gene can be varied according to patient response or terminated in the event of a severe adverse effect. Therefore, the current generation of gene promoters used in gene therapy clinical studies do not meet all the requirements to enable safe and effective cell and gene therapy. It is imperative that the industry makes use of recent advances in understanding how genes are regulated in the human genome to build more rationally-designed therapeutic expression cassettes.

Making Use of Big Data

The sequencing of the human genome at the turn of the new millennium heralded a new era of biology. We not only learned the location and identity of new gene coding sequences, but that the human genome is a blueprint of gene regulation, where the mechanisms that control the timing and restricted pattern of gene expression are embedded within the sequence itself. Access to the complete sequence of the genome also facilitated a huge surge in bioinformatics, where scientists began the monumental task of annotating the genome. Of importance, databases such as TRANSFAC and JASPAR were established to map out how transcription factors orchestrate the complex networks that control global gene expression (1-2). This allowed researchers to establish exactly where the gene transcription machinery binds to the genome, which subsequently enabled the identification and characterisation of endogenous gene promoter sequences. These databases eventually laid the foundations for more comprehensive efforts led by large consortia to map entire genomic regulatory regions.

Encyclopaedia of DNA Elements (ENCODE)

ENCODE consists of employed methods comprising many sequence-based studies to map functional elements across the human genome (3). The elements mapped (and approaches used) include RNA-transcribed regions (RNA-seq, cap analysis of gene expression [CAGE], RNA-PET and manual annotation), protein-coding regions (mass spectrometry), transcription-factor-binding sites (ChIP-seq and DNase-seq), chromatin structure (DNase-seq, FAIRE-seq, histone ChIP-seq and MNase-seq) and DNA methylation sites (RRBS assay).

Functional Annotation of the Mammalian Genome (FANTOM)

By applying short RNA-sequencing and CAGE, this consortium generated expression profiles of mature miRNAs and their primary transcripts across a large collection of primary cell types (4). CAGE also allowed the mapping of endogenous promoters and enhancers that control expression of RNA polymerase 2-derived mRNA.

NIH Roadmap Epigenomics Mapping Consortium

This is a public resource of human epigenomic data led by a consortium to leverage experimental pipelines built around next-generation sequencing technologies to map DNA methylation, histone modifications, chromatin accessibility and small RNA transcripts in stem cells and primary *ex vivo* tissues selected to represent the normal counterparts of tissues and organ systems frequently involved in human disease (5).

The Genotype-Tissues Expression Consortium (GTEx)

GTEx provides to the scientific community a resource with which to study human gene expression and regulation and its relationship to genetic variation (6). This project collects and analyses multiple human tissues from donors who are also densely genotyped to assess genetic variation within their genomes. By analysing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as expression quantitative trait loci.

Each of these resources collate from diverse sources of global gene expression studies and high-throughput genomic analysis methodologies to establish how unique transcription networks determine the precise molecular phenotype of a cell (7). Consequently, the industry now has a deeper understanding of the actual sequences within the genome that control the expression of individual genes and, through participation cell-type specific genetic programmes, ultimately determine a cell's fate. Collating, interpreting and employing data generated from these functional genomics resources is no simple task. The data are complex and difficult to interpret and are often used in deciphering transcription networks to understand cellular differentiation and developmental processes.

To identify which transcriptional elements are important in regulating cell type-specific genes in selected environmental conditions, bioinformatics scripts and machine learning algorithms have been developed that integrate data from a diverse array of genomics sources. These algorithms allow identification of putative enhancer regions that can be further analysed either by a) human researchers by using customised genome browsers or b) by machine-learning protocols that automatically recognise DNA modification patterns in the selected sequence that have hallmarks of enhancer function. This dual approach is used to essentially rank cell type-specific enhancers in order of likely biological relevance. Typically, when developing machine algorithms designed to rank putative enhancer sequences, data is analysed on histone modifications, DNA methylation, transcription start site analysis, nucleosome positioning and a vast array of other functional genomics techniques that are widely employed to interpret how the human genome is functionally transcribed.

Future Challenge

This new wealth of genomic information gleaned from function genomics and bioinformatics experimental analyses has generated a large collection of functional genetic parts. Many companies have looked to exploit this information to create smarter 'synthetic' promoters designed to switch on genes in any given environment or in response to any change of that environment.

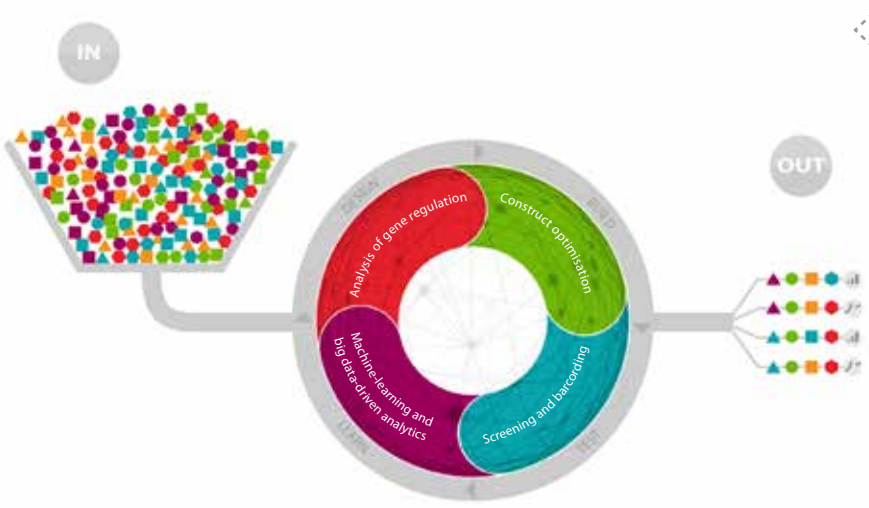

Figure 1: A platform for the creation of synthetic promoters. Genetic parts are identified as INPUT from the human genome by analysing functional genomics datasets using a series of dedicated bioinformatics scripts. An engineering biology 'design, build, test, learn' approach is adapted to build complex libraries of synthetic promoters comprising these parts, which are screened using high-throughput complex screens. The OUTPUT are synthetic promoter sequences that comprise novel combinations of natural gene regulatory sequences derived from the genome

To achieve this, engineering biology techniques are being employed where promoter, enhancer and terminator parts are identified, which are standardised to be used 'off-the-shelf' to build novel therapeutic gene expression cassettes (see Figure 1). Once identified, novel library screening technology can be employed to resolve the precise enhancer sequences that control transcription, which then can be used as parts to build synthetic promoters using a rational engineering biology approach. The result is a novel promoter that tightly controls gene expression in the particular cell type and condition of interest and comprises a sequence constituting a novel combination of enhancer elements that does not exist in nature and can thus be patented.

One exciting application of this approach is the enhancement of gene- and cell-based therapies, so that therapeutic genes are put under the control of novel combinations of genome-derived enhancer regions, designed to express precise therapeutic levels of a protein in a controllable manner, specifically in the desired cell type and over a longer period, thus addressing some of the issues described earlier.

References

1. Visit: www.gene-regulation.com/pub/databases.html
2. Visit: jaspar.genereg.net
3. Visit: www.genome.gov/encode
4. Visit: fantom.gsc.riken.jp
5. Visit: www.roadmapepigenomics.org
6. Visit: www.gtexportal.org/home
7. Suryamohan K and Halfon MS, Identifying transcriptional cis-regulatory modules in animal genomes, *Wiley Interdiscip Rev Dev Biol* 4(2): pp59-84, 2015

Dr Michael L Roberts studied biochemistry at the University of Glasgow, UK, and completed his PhD at the University of Cambridge, UK, where he employed viral vectors to study plasticity in the peripheral nervous system. He then proceeded to a post-doctoral position at Royal Holloway, University of London, where he worked on developing novel gene therapies for neuromuscular disorders. In 2002, Michael moved to Greece on a Marie Curie fellowship to set up a functional genomics facility at the National Hellenic Research Foundation. After spending five years running gene therapy R&D activities for a small US biotech firm, Regulon, he moved back to Edinburgh in 2010 to establish Synpromics. Michael currently serves as the company's Chief Scientific Officer.
Email: michael.roberts@synpromics.com