The Path to Discovery: Danuta Jeziorska

Shining a light on the dark genome

Over half a century ago the helical structure of DNA was identified and 40 years later the Human Genome Project was born. Inevitably, there was high interest across the healthcare and scientific sectors – what could be more powerful than having access to the full instruction manual to our cells? It took 13 years to generate the first sequence and its announcement in 2000 anticipated a new era of drug discovery and precision medicine.

Of course, we did not foresee the complexities of that instruction manual – it took us by surprise that only 2% of the 3 billion base pairs that make up our genome encoded for proteins – equating to approximately 20,000 genes. Whilst the potential of these gene coding regions began to be exploited across the pharmaceutical sector, the remaining 98% of the genome, the 'dark matter' or 'dark genome', was not well understood and considered by some as 'junk' at that time.

To better define the genetic origin of diseases, an avalanche of studies was initiated looking for associations between the millions of genetic variations in the genome and known diseases. Starting with age-related macular degeneration, the genome wide association studies (GWAS) associated nearly 5,000 diseases and traits and identified thousands of unique genetic loci.[1] Then came the second big surprise: only 5% of the genetic variants were located within protein coding genes. The studies had clearly worked as around 70 of known licensed drug targets were 're-discovered' via variants. These included coding variants for PPARG (peroxisome proliferator activated receptor gamma), the target of the thiazolidinediones for type 2 diabetes.[2]

In an industry where success rates are low and investment is high, there was predictably a strong desire to unlock the potential of the 95% of disease-associated genetic changes located within the dark genome, particularly as we knew that genetic evidence improved successful drug development by more than two-fold.[3] Due to technological and scientific advancements, we have elucidated that the dark genome is not junk but rather has a critical regulatory function to play. It controls cell type specific gene expression, acting like an instruction processor and turning genes on and off at the right time and level, enabling our genetic code to be translated into the hundreds of cell types that exist. It is not hard to understand how dysregulation of this mechanism can cause evolutionary modification or human disease.

Indeed, publications in 2016 studying the genetic evolution of snakes indicated that mutations in the dark genome, located near to a gene essential for limb formation, have kept that gene from ever being turned on, explaining the snake's legless body.[4] It was a simple leap of logic to assume that all dark genome variants were linked to a neighbouring gene. However, results pursuing that approach were disappointing. In addition, computational approaches had limited impact as DNA was treated as a linear structure and methods could not decipher the apparently distal associations. However, discovery of a 3D genome analysis method[5] dramatically enhanced our understanding. Just as proteins fold to make 3D structures, so does DNA. When considered in three dimensions, regulatory elements in the dark genome come into close proximity to activate apparently distal genes. We had the ability understand the complexities of the relationships between these dark genome regulators and the small percentage of gene-coding regions.

At Nucleome, a company I founded in 2019 with Prof Jim Hughes and Prof James Davies from the University of Oxford, we have a platform of technologies that allow us to leverage genetic variations from the dark genome to identify novel and safer drug targets, de-risk and repurpose existing targets, as well as help us better understand complex diseases.

World leading position in 3D genome analysis

We have a world leading position in 3D genome analysis for interpreting genetic changes in the dark genome. Our methods[6,7] offer significant benefits over other methods currently available including the ability to work with primary cells and patient derived cells; and incredible accuracy – an average of 256 base pair (bp) resolution[6] and unprecedented 1 bp precision for our recently invented Micro-Capture-C (MCC) method (published in *Nature* last month[7] and described in the Box in more detail). This is important as disease-associated variants are often only single bp changes within regulatory elements of 50 – 1000bp in size, which are often surrounded by other functional elements and multiple genes. Therefore, to make accurate associations, a high degree of resolution at the DNA level is required.

We have already investigated over 3.5 million genetic variants, across more than 20 different immune cell types using our machine learning and computational genomics tools. This allowed us to understand which variants are affecting regulatory elements that function in the dark genome and in which cell types; as well as to explore the links between cell types and more than 1,000 diseases in an agnostic manner. To facilitate scalable and smart variant prioritisation and exploration of this data, from the perspectives of cell types, phenotypes, variants and GWAS studies, Nucleome has built 'Lantern', a proprietary database. Currently, we are associating genes with prioritised variants to link them with diseases.

Once the linkage of a genetic variant and gene activity has been made, these relationships and their effects will be functionally validated in cells. Typically, this validation is approached via genetic engineering methods such as CRISPR. However, this presents a large bottle neck when considering thousands of variants. Nucleome uses

©2021 Evernow Publishing Ltd

a proprietary method that allows us to do this at scale, in primary cells without genetic engineering. This scalability offers us the ability to map disease affected pathways and allows us to explore individual targets or the cell as a system. As Nucleome's platform can be applied to multiple cell types, this has the potential to discover novel high-quality targets with corresponding biomarkers across multiple diseases at the level of DNA, genes, and pathways.

Our vision is to decode how the genome functions in health and disease, opening the door to a new generation of therapeutic discoveries and making the human genome actionable for the benefit of patients. Initially we are focusing on lymphocytes and immune driven diseases with the ambition to build a pipeline of assets.

With a dedicated team who are passionate about making a difference to the lives of patients with unmet medical need, we are working towards unlocking the genome for precision medicine and the future envisioned by former US president Bill Clinton who said on the day of the announcement of the Human Genome Project: *"It is now conceivable that our children's children will know the term cancer as only a constellation of stars."*

References:

1. Loos RJF., *Nature Communications* 11: 5900 (2020)
2. Finan C., et al., *Sci Transl Med*. 9: 383 (2017)
3. Nelson MR., et al., *Nature Genetics*, 47: 856-869 (2015)
4. Kvon EZ., et al., *Cell*, 167: 633-642 (2016)
5. Dekker J., et al., *Science* 295, 5558:1306-1311 (2002)
6. Davies JOJ., et al., *Nature Methods*, 14: 125-134 (2017)
7. Hua P., et al., *Nature*, 595: 125-129 (2021)
8. The ENCODE Project Consortium et al., *Nature* 583: 699-710 (2020)

This article was written by Dr Danuta Jeziorska, chief executive officer and founder of Nucleome Therapeutics Ltd.

3D genome architecture at single base pair precision

Until the 1950s, the structure of DNA was a mystery. We knew from Levene and others that DNA was composed of nucleotides – adenine (A), thymine (T), guanine (G) or cytosine (C) – that were covalently linked into chains. We didn't know how a two metre length of DNA could be packaged within the nucleus of almost every cell. Following pivotal studies by Watson and Crick, inspired by Franklin's Photo 51 and Chagraff's nucleotide ratios, the DNA double helix was revealed. But this was only the first step in understanding how DNA is organised. We now know that DNA also undergoes multiple packaging strategies: DNA wraps around histone proteins to form tight loops called nucleosomes and these coil and stack together to form fibres called chromatin. We also know that these structures and loops (in the so-called 'dark genome') play a key role in regulating gene activation in a cell type specific manner. However, deciphering which regulatory elements (we have around 810,000 of them[8]), affect which genes has been one of the key stumbling blocks to fully understanding the dark genome's effects in disease.

The chromosome conformation capture (3C) method[5], that detects the frequencies of interactions between DNA sequences, has revolutionised the field over the last decade. Until recently however, this and similar approaches have not achieved the level of resolution that has the potential to unlock the dark genome's secrets with base pair precision. However, Micro-Capture-C (MCC)[7] achieves on average 256 – 1000x higher resolution, reaching this holy grail of single base pair accuracy. So how does this technology work?

The process starts with chemical crosslinking to fix interacting DNA fragments. This crosslinked DNA is cut using the enzyme micrococcal nuclease (MNase), which cuts mainly in a random fashion. This is in contrast to sequence specific enzymes, commonly used in other methods. Next DNA is re-connected using another type of enzyme, a ligase. These steps are done in intact permeabilized cells, instead of solubilised chromatin or purified nuclei; thus minimising disruption of the chromatin architecture. The DNA is then de-crosslinked and extracted. As interacting DNA is now linked together (reflecting the 3D rather than linear structure) genome-wide 3D interaction profiles can be generated at scale. This method is being used at Nucleome to decipher the dark genome and provide the largely missing links for interpreting disease linked genetic variants that affect its function.

MedNous >>>

Reliable and up-to-date

MedNous
Opportunities in European Medical Innovation
Business Strategy
How artificial intelligence is changing drug discovery
Research Strategy
Novel immunotherapy arises from collaboration
Oncology Review
Long-term strategies for immuno-oncology
Licensing and Collaboration Briefs
Plus
Subscribe NOW

Visit www.mednous.com >>>>>>>>>>>>