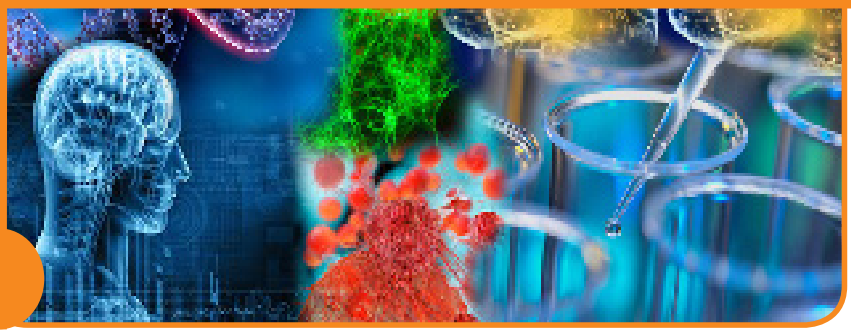
Advances in Genetics and Stem Cell Research open New Frontiers

Dr Aruna Dharshan De Silva
Mr Harindra Sathkumara



Humans are a complex life form made up of trillions (~1012) of cells, the basic building block of all living things. We all start this amazing journey of becoming a full grown human from a single cell stage called a 'zygote' which results from fertilization of two gametes, one from each parent. But how does a simple single-celled zygote give rise to a multicellular organism with all of its specialized organs? The basis of growth and development of all biological systems is cellular growth, cell division, and cell differentiation. The perfect balance between these exquisitely choreographed sequential events is the key for human development, which entails growth from a single cell to an adult human being. A group of cells called 'stem cells' play a key role in the process of human growth to become an adult human from the zygote stage. So what makes the stem cells differ from other types of cells? Stem cells have a remarkable potential to self-replicate and to differentiate into many specialized cell types throughout the embryonic development and growth, rendering them one of the most fascinating areas to explore (Figure 1). But what truly determines who we are, is the genetic code (DNA) that is embedded in our cells which contains all of our genes. DNA was not known for carrying information in cells until 1944 when it was discovered by Avery, McLeod and McCarty as being responsible for transferring "genetic material". We now understand how we inherit

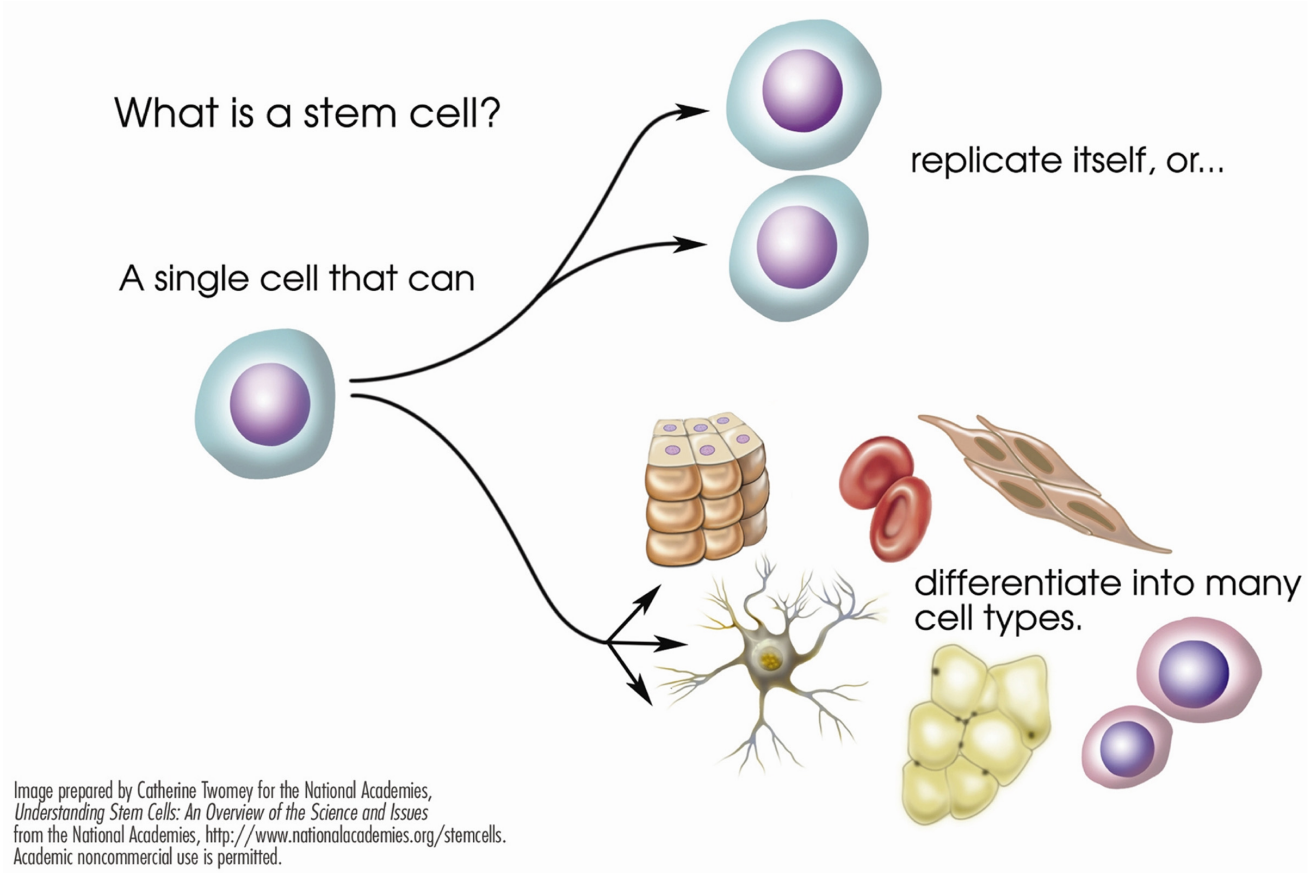


Image prepared by Catherine Twomey for the National Academies, *Understanding Stem Cells: An Overview of the Science and Issues* from the National Academies, http://www.nationalacademies.org/stemcells. Academic noncommercial use is permitted.

Fig 1: What is a stem cell?

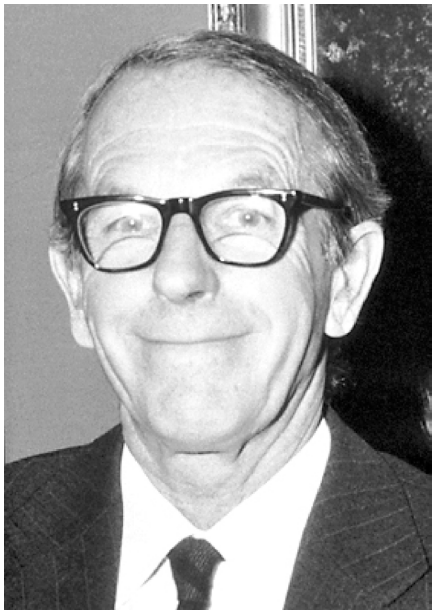


Fig 2: Frederick Sanger (1918-2013)

the characteristics or traits from our parents. This article outlines how advances in our knowledge of genomes and genetic sequencing have opened new frontiers in the areas of genetics and stem cell research.

Our genetic information is stored in the form of four bases (i.e. adenine (A), thymine (T), cytosine (C), and guanine (G)), arranged in different ways and in different lengths. This order spells out the exact instructions required to create a particular organism with its own unique traits. The human genome occupies about three billion (~3x109) base pairs and surprisingly only about 1.5-2% codes for a protein while a large portion of the genome consists of non-coding RNA genes, regulatory sequences, introns, and noncoding DNA. Furthermore all the cells in the human body contain the exactly same DNA or genome. If all the genes are directly expressed as they are, then all the trillions of cells in our body should be morphologically and functionally identical. Then what makes a skin cell different from a red blood cell or a neuron? The answer lies on how each cell deploys its genome, or in other words, how the combination of genes that are turned on (expressed) or turned off (repressed) determines its fate. With ever increasing enigmas, scientists began to realize that cracking the human genetic code would be the only way to decipher the unsolved mysteries behind the complex human living organism. It was believed that learning to decipher the DNA code may lead to revolutionary new ways to diagnose, treat, and someday prevent the thousands of disorders that affect us.

The Human Genome Project (HGP) revealed the initial draft of mankind's DNA sequence in 2001. The primary goal of HGP was to figure out the sequence of billions of letters that make up the human genome. . It was not an easy task deploying the effort of hundreds of scientists across dozens of countries, and costing over three billion dollars (US$ ~3x109). It took thirteen years (1990 to 2003) to sequence and map all the genes in this effort. Development of the chain termination method (commonly referred to as the Sanger method) by Sanger and colleagues in 1970s was considered the gold standard for nucleic acid sequencing for the subsequent two and a half decades, that established the groundwork for decades of sequence-driven research that followed. The most notable use of Sanger technology in DNA sequencing was the HGP. Frederick Sanger (Figure 2) won his

What NGS does?
• NGS provides a much cheaper and higher-throughput alternative to sequencing DNA than traditional Sanger sequencing. Whole small genomes can now be sequenced in a day. • High-throughput sequencing of the human genome facilitates the discovery of genes and regulatory elements associated with disease. • Targeted sequencing allows the identification of disease-causing mutations for diagnosis of pathological conditions. • RNA-seq can provide information on the entire transcriptome of a sample in a single analysis without requiring previous knowledge of the genetic sequence of an organism. This technique offers a strong alternative to the use of microarrays in gene expression studies.
And limitations
• NGS, although much less costly in time and money in comparison to first-generation sequencing, the initial costs are too high for many labs. NGS platforms can cost more than $100,000 in start-up costs, and individual sequencing reactions can cost upward of $1,000 per genome. • Inaccurate sequencing of homopolymer regions (spans of repeating nucleotides) on certain NGS platforms, including the Ion Torrent PGM, and short-sequencing read lengths (on average 200–500 nucleotides) can lead to sequence errors. • Data analysis can be time-consuming and may require special knowledge of bioinformatics to garner accurate information from sequence data.

Fig 3: Popular NGS platforms currently available at the market

second Nobel Prize in 1980 for his contribution in the determination of base sequences in nucleic acids. An educational animation on how Sanger method works in relatively modern automated sequencers can be found at Wellcome Trust Sanger Institute's website. http://www.wellcome.ac.uk/Education-resources/Education-and-learning/Resources/Animation/WTDV026689.htm

Sequencing technology has evolved at a very fast pace over the past ten years, and the Sanger method has now been partially replaced by several next generation sequence (NGS) platforms that offer dramatic increases in cost-effective sequence throughput, which have had a tremendous impact on genomic research. The most remarkable feature of NGS technologies is its capacity to perform massively parallel sequencing, during which millions of fragments of DNA from a single sample are sequenced simultaneously allowing an entire genome to be sequenced in a very short time in contrast to decades with Sanger. Several NGS technologies (2nd generation) have been developed using varied approaches since the HGP, each with its own unique strengths and shortcomings (Figure 3).

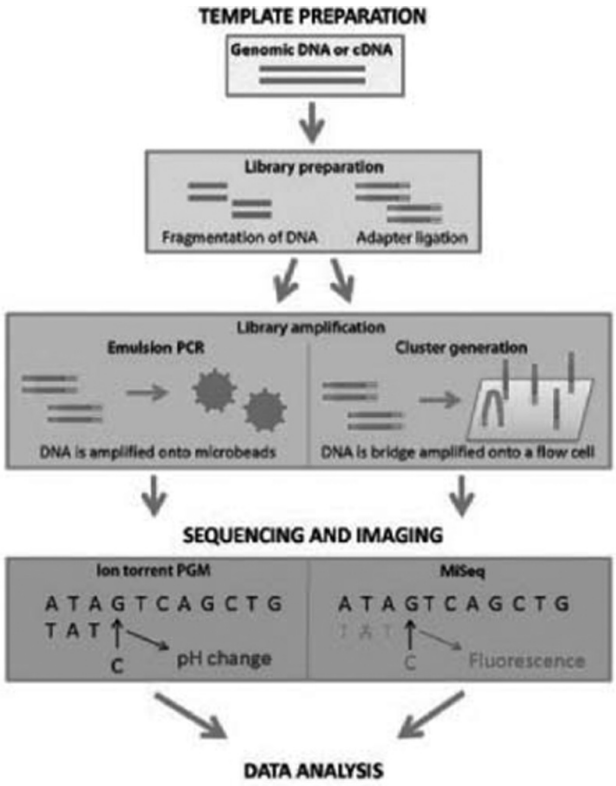


Figure 4: Basic workflow for Ion Torrent (left) and illumina MiSeq (right) sequencing technologies

Although each NGS platform is unique in how sequencing is accomplished, the Ion Torrent PGM and the Illumina MiSeq have a similar base methodology that includes template preparation, sequencing and imaging, and data analysis (Figure 4). Both platforms are more popular and currently available in Sri Lanka.

Table 1 shows some characteristic features of high-end 2nd generation sequencing platforms and more recent bench-top platforms. While Ion Torrent relies on detection of pH changes that occur during the incorporation of nucleotide into the growing DNA strand, in the MiSeq, template molecules are sequenced in a massively parallel fashion using a DNA sequencing by synthesis (SBS) approach and relies on the detection of fluorescence generated by the incorporation of fluorescently labeled nucleotides into the growing strand of DNA (Life Technologies and Illumina's company websites have more info).

Regardless of the different sequencing approaches, all NGS platforms have enabled scientists to

Table 1: Characteristic features of 2nd generation sequencing platforms 5

High-end sequencing- Platform†	Sequencing chemistry	Read lengths/ through put	Run time	Template prep	Application
Roche 454 -Titanium FLX	Pyrosequencing	400 bp 400 Mb/run	10 hours	Emulsion PCR	Denovo WGS of microbes, pathogen discovery, Exome seq
Illumina/Solexa -HiSeq 2000	Reversible terminator chemistry	2×100bp 600 GB/ run (dual cell)	11.5 days	Solid-phase	Human WGS, exome seq, RNA-seq, Methylation
ABI/LifeTechnology-SOLiD 5550XL	Sequencing by ligation	2×60bp 15 GB/day	8 days	Emulsion PCR	Human WGS, exome seq, RNA-seq, Methylation
HelicosBiotechnologies	Reversible Terminator chemistry	25-55 bp 28 GB/run (avg)	>1 GB/hour	Single molecule	Human WGS, exome seq, RNA-seq, Methylation
Roche 454- GS Junior	Pyrosequencing	400 bp 50 Mb/run	10 hours	Emulsion PCR	Denovo WGS of microbes, pathogen discovery, Exome seq
Illumina/Solexa- MiSeq	Reversible terminator chemistry	2×150bp 1.0-1.4 Gb	26 hours	Solid-phase	Microbial discovery, Exome seq, Targeted capture
ABI/ Lifetechnology- Iontorrent	H+ Ion sensitive transistor	320 Mb/run	8 hours*	Emulsion PCR	Microbial discovery, Exome seq, Targeted capture

*Sample preparation – 6 hours, sequencing time – 2 hours, †Data shown here represent the highest figures currently available on the company website and is highly likely to change by the time this article is published

not only sequence a whole genome of a human in a week, but also to appreciate the new aspects of the sequencing such as advances in the characterization and quantification of transcriptomes using RNA-seq. Most importantly previously unknown coding and non-coding RNA species, particularly small RNAs, including microRNAs (miRNA) and small interfering RNAs (siRNA) which seem to play a pivotal role in gene expression regulation in patients with certain disease conditions such as cancers. Currently, scientists are hard at work interpreting the genome. While not every difference is consequential, some of these differences are responsible for how we look, how we act and even how likely we are to get sick or to respond to specific medicines. Better understanding on how disparities between our genomes account for these differences is sure to change the way we think. The applications of NGS seem almost endless, allowing for rapid advances in many novel fields related to the biological sciences including analysis of epigenetic modifications and metagenomics.

With all the sequencing information in hand, scientists were not keen on assessing the impact of genes of interest in an *in vitro* disease model No 5, preferably an animal model since human is the mammal we cannot use for genetic experiments for obvious ethical reasons. Interestingly, mice naturally develop conditions that mimic human disease. Recent development in molecular biology and stem cell biology has allowed scientists to induce diseases in mice by manipulating its genome or to create custom-made mice through a process called site-directed mutagenesis in mouse embryonic stem (ES) cells. There are 2 major mice genetic mutation models, transgenic and knockout. In transgenic mice, an additional piece of DNA is inserted and depends on the expected goals of the experiment; mice will exhibit different expression levels of proteins, such as GFP (Figure 5). Functional aspects of a protein can only be learnt by deleting the gene that codes the protein or the deletion

Fig 5: Transgenic mice expressing Green Fluorescence Protein (GFP)8

Fig 6 : Mario R. Capecchi

of a functional domain of the protein. This can be achieved by using gene knockout mice. Dr. Mario R. Capecchi (Figure 6) of the University of Utah won the 2007 Nobel Prize in Physiology or Medicine for his pioneering work on knockout mice technology. This technology has given scientists worldwide; an amazing set of tools to make important discoveries about how defective genes could lead to human diseases.

Mutations (either inherited or sporadic) in the coding sequence (CDS) and/or defective gene expression mechanisms have been found to cause a spectrum of diseases in humans, ranging from obesity to cancer. Bone marrow transplantation (BMT) has now been considered the most effective way of treating a number of blood cancers and primary immunodeficiency disorders (i.e. leukemia, severe combined immunodeficiency). The driving force behind BMT is the hematopoietic stem cells. BMT replaces damaged stem cells with healthy ones, which give rise to blood-forming cells, restoring the normal function.

Stem cell research and gene therapy studies hold great promise of bringing new ways of identifying potential disease targets and developing novel therapies. There are 2 types of stem cells, embryonic stem(ES) cells and adult stem cells. While ES cells can generate all types of cells, adult stem cells can differentiate to yield the major specialized cell types of the specific tissue or organ. For instance, adult stem cells in heart can only give rise to heart tissues. Ongoing stem cell research has passed major milestones over the last few years and is rapidly moving towards achieving the ultimate goal of personalized medicine. Some research groups have already taken initiative to test possibilities of making specialized disease cells using induced pluripotent stem cells (iPSCs) thus the efficacy and safety of various existing and novel drugs could be tested on the cultured cells instead of directly on the patient. This will allow us for the first time to figure out what exactly has gone wrong in rare genetic diseases such as Riley-Day syndrome(disorder of the autonomic nervous system). Regenerative medicine is also moving towards a day when we can repair and replace damaged tissues like we have seen in many futuristic science fiction movies. Scientists have already been able to successfully grow whole functional organs not only inside a living organism (a thymus inside a mice) but also in a research lab (a lab-grown rat kidney is shown in Figure 6). Gene therapy is rapidly moving forward with advanced knowledge in genetics and breakthroughs in the understanding of stem cell functions. In time, we will be able to seek treatments such as insulin-secreting pancreatic cells, bone cells capable of healing breaks and defects, and eye and ear cells to restore vision and hearing.

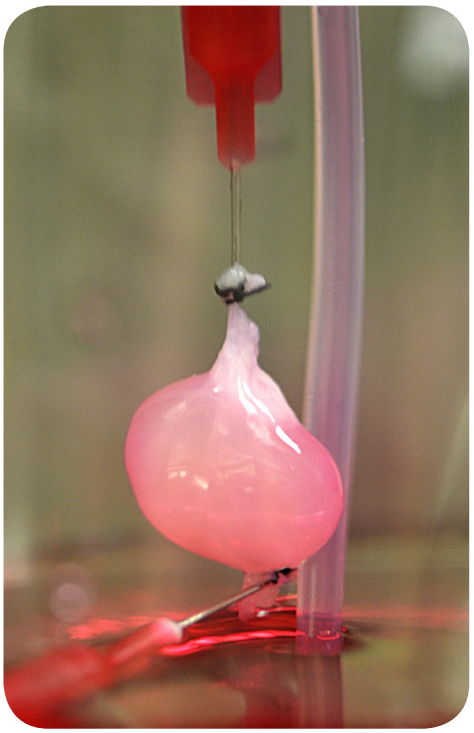


Fig 7 : Lab-grown rat kidney 10

Dr Aruna Dharshana De Silva
Director
Senior Scientist
Genetech Research Institute
T/P : 4940349

Mr Harindra Sathkumara
Scientific Officer
Genetech Research Institute
T/P : 4940349