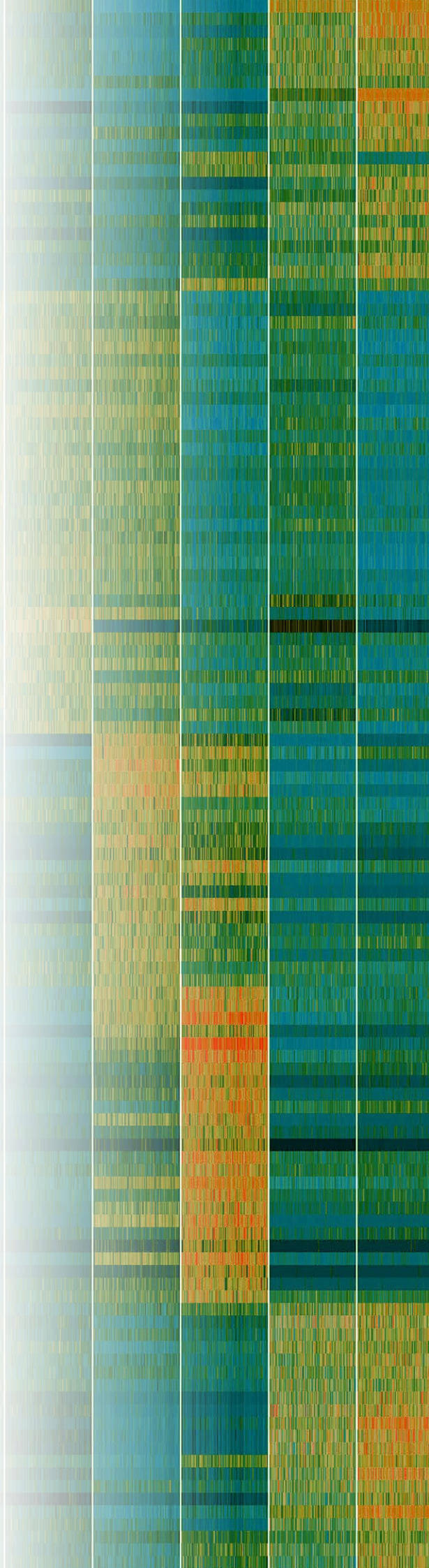
White paper

Exploring the landscape of single cell RNA sequencing



Humanizing Genomics
macrogen EUROPE

Exploring the landscape of single cell RNA sequencing

The last decade has clearly seen the maturation of single cell RNA transcriptomics. From its early beginnings as a laborious, low-throughput technique to a trusted, well-recognized approach – single cell RNA-seq is now expanding its reach to an ever-increasing number of scientific fields. And with a fast succession of innovations in single cell workflows, this expansion is showing no signs of slowing down.

The 'why' of single cell RNA-seq

Single cell RNA-seq finds its roots in a range of more established techniques such as bulk RNA sequencing, microarrays, northern blots and quantitative PCR. These technologies are well established and have been used for many years to study RNA. Data from these assays has provided vast number of insights into the transcriptome of many species.

However, the key limitation of bulk RNA sequencing arises when scientists are interested in differences in the gene expression profiles of cells within a sample. In bulk sequencing, the RNA from all cells in a sample is combined to form a single sequencing library, which means that the analysis of the sequencing data will paint a picture of the average expression profile of a cell. Although this data is extremely valuable for a lot of different genomics applications, there are also many situations where the important information cannot be extracted by looking at averages alone.

An example where average expression data gives incomplete information is a sample that contains two or more subpopulations – a feature often found in tumor biopsies. A cell population might appear homogeneous when isolated or when analyzed using other techniques – but might consist of distinct subpopulations when gene expression profiles are taken into account.

Another common example is the presence of rare cell types in a sample, for example in a liquid biopsy. The expression profile of these cells might be very different from the main population, and their role could be highly significant, but in bulk RNA sequencing their presence or activity can easily be hidden due to their low prevalence (figure 1).

With increasing awareness, knowledge and availability of single cell RNA-seq technology it has become a viable option for an increasing number of applications. These applications can now benefit from the increased precision and comprehensive data that single cell technologies have to offer.

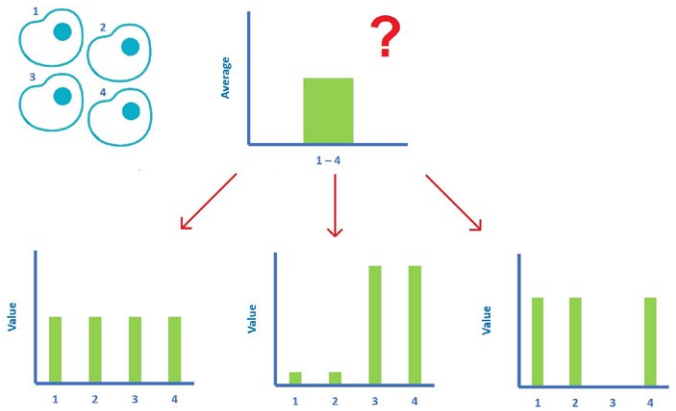


Figure 1.
Schematic representation of one of the limitations of bulk RNA sequencing. The same value for the average expression level of a gene can be obtained from a homogeneous sample (left), a sample containing two subpopulations (middle) and a sample containing a rare cell type with an abnormal expression level (right).

Single cell RNA-seq making a difference

The unique ability of single cell RNA-seq to detect transcriptome variations within samples that might otherwise appear homogeneous makes it well suited to a range of scientific studies. These fields increasingly rely on single cell data to obtain a more complete picture of gene expression patterns.

Immunology

Immunology is one of the fields where cell heterogeneity holds critical information for many important biological questions[1]. Both pathogens and the immune cells created to stop them cannot be analyzed adequately when seen as a single, uniform group. A single cell approach is needed to understand the underlying genomic mechanisms in many immunology-related fields of research (figure 2).

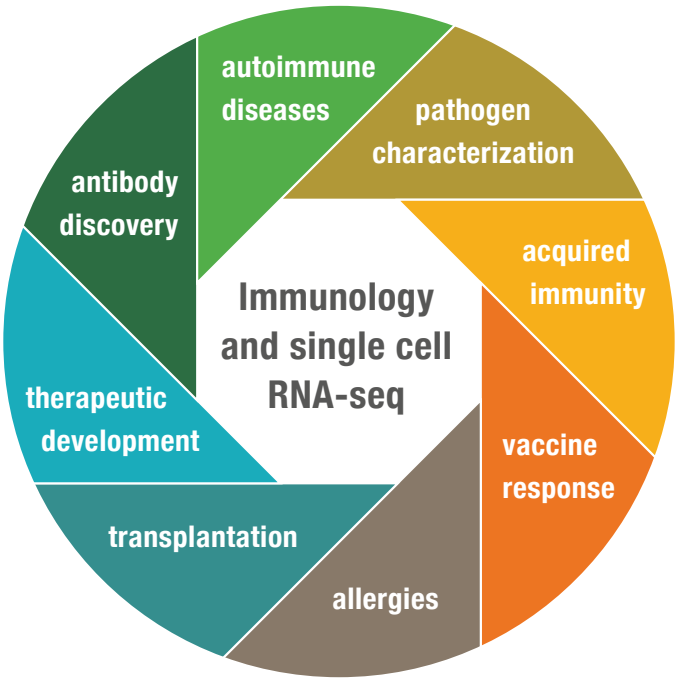


Figure 2.
Applications of single cell RNA-seq in immunology.

When it comes to pathogen characterization, single cell RNA-seq can play a role in distinguishing between closely related pathogens, as well as between different transcriptional stages of the same pathogen. Mutations that affect a pathogen's function – and the disease states it causes – can be found by studying the transcriptome, but single cell resolution is instrumental in studying these mutations in detail. In addition, many pathogens go through different transcriptional stages during their life cycle. These changes are also often hard to study when using bulk RNA sequencing alone[2].

Tasked with fighting the these ever-evolving microbes, the cells of the immune system show an equally complex transcriptome. The human adaptive immune system is critically dependent on the expression of a diverse repertoire of antigen receptors – immunoglobulins and T cell receptors – by B and T cells[3]. One key way in which the immune system achieves this diversity is by V(D)J recombination.

The V(D)J region refers to a segment in several antigen receptor genes that undergoes a high level of rearrangement during lymphocyte maturation. In their original state, these segments consist of series of variable (V), (diversity (D)) and joining (J) elements – which are separated by recombination signal sequences (RSSs) and other control elements. During maturation, protein complexes create double stranded breaks in the DNA, which are subsequently repaired by non-homologous end joining – leading to high receptor diversity in mature B and T cell populations.

Single cell RNA-seq enables researchers to explore and visualize T and B cell diversity in great detail (figure 3)[4]. In addition, a combined analysis of gene expression and V(D)J region interrogation can give relative abundances of the different types of immune cell in a sample – as well as variation within these subpopulations.

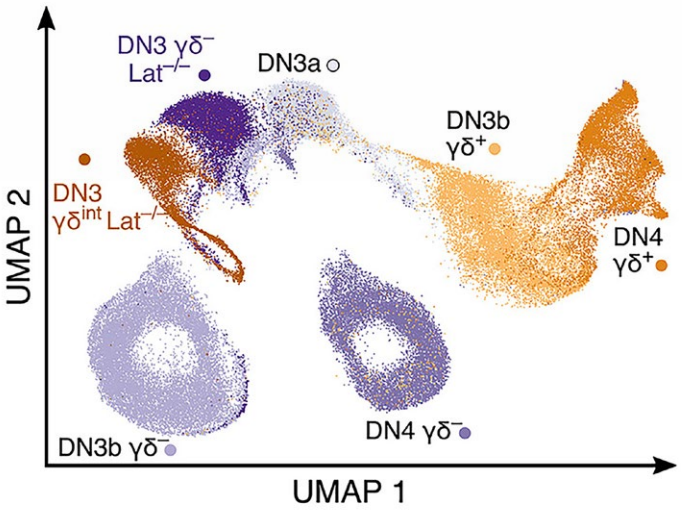


Figure 3.
By using dimension reduction, for example using the uniform manifold approximation and projection (UMAP) method, 2D plots can be created to visualize key features in single cell RNA-seq data. Figure reproduced, with permission, from S Scaramuzzino, D Potier et al.[5]

As an example, a 2022 study by He et al. used this combination of single cell gene expression and single cell VDJ sequencing to analyze bone marrow biopsies of patients with multiple myeloma[6]. This malignancy is characterized by high levels of genetic heterogeneity, which results in a high risk of relapse after treatment.

With the aim of finding new biomarkers for effective, targeted treatments, the researchers compared single cell data for CD138+ malignant myeloma cells of both newly diagnosed and refractory/relapsed multiple myeloma. Using the Cell Ranger and Seurat packages, they analyzed malignant clonotypes, carried out cell type annotation and identified disease status-specific genes.

Using single cell genomic data the researchers discovered that the myeloma cells were dominated by a major clone, but also identified three clusters with novel differentially expressed genes (DEGs) that were exclusively higher in relapsed samples. In addition, they identified several attractive biomarker candidates that were associated with disease progression and relapse, including SMAD1 and STMN1.

Oncology

Research into cancer risk, development and progression is another key area in which single cell events can have major implications. Many cancers develop through an accumulation of mutations – caused by genomic instability – which means that subpopulations with unique expression profiles are common. This heterogeneity is often considered to be the main driver of metastasis and resistance to therapy[7].

In a recent study by Wouters et al. for example, researchers carried out detailed single cell transcriptomic analysis of melanoma cell cultures to investigate phenotype switching and the presence of intermittent states existing in between the known phenotypes[8]. Melanoma cells are known to exist in both a melanocytic and a mesenchymal-like transcriptomic state, and are able to switch between them.

By analyzing each culture at the single cell level, the researchers managed to prove conclusively that a distinct intermittent state exists, and that earlier evidence for the existence of this state was not the result of mixed populations of the two known states. They also showed that the phenotype switching is not a stochastic process, but instead is closely regulated – and they also identified five key genes involved in regulation.

Rare cell types is another population feature that can be highly relevant to many oncology researchers. These can be, for example, novel mutations within a sample of tumor cells, but also circulating tumor cells (CTCs) in a blood sample. Single cell transcriptome data can help researchers find these rare cell types and study them in great detail.

These same benefits of single cell transcriptomics are also showing their value in immuno-oncology, where they help to shine light on the complex role between a tumor, the tumor microenvironment, and the immune system. Advanced, targeted cancer treatments such as immune checkpoint therapy have a large potential for growth in the coming years, but the development of safe and effective treatments is critically dependent on detailed characterization of single cells.

Getting started with single cell RNA-seq

Despite the obvious benefits that single cell transcriptome data can offer, many will be deterred by the complexities of setting up a single cell RNA-seq workflow. And not without reason. Starting out in the world of single cell transcriptomics might bring challenges not found in bulk RNA sequencing and other molecular assays – and that need careful consideration before getting started.

The first challenge many researchers will face is sample preparation. The quality of single cell sequencing data is critically dependent on the starting materials used, which means that generating a high-quality single cell suspension with a high cell viability will be key to getting the best results. Methods for generating good single cell suspensions, however, vary considerably depending of the tissue type, and sample-specific optimization is often required.

Secondly, there is the choice of scale of the study. Scale in this case will be determined by the number samples, cells and sequencing reads of the study. All of these will dependent on the scientific hypothesis, practical limitations and budget – and will in turn determine the optimal workflow for the study. In recent years, many different approaches for single cell RNA-seq have been published, validated and commercialized, and the optimal choice will always – among other factors – depend on throughput[9].

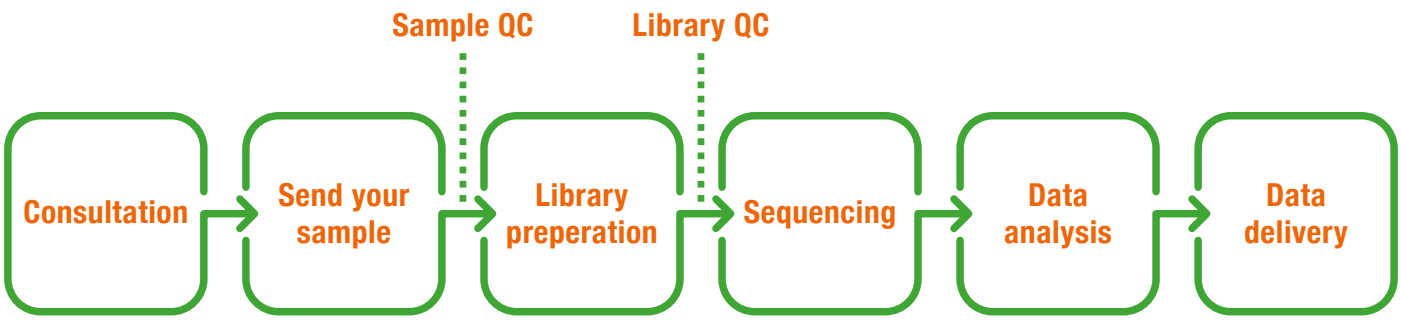
Finally, data analysis will be an area where there are many choices to be made, not just about addressing a scientific hypothesis, but also for reducing errors in the data. Examples of the latter include datasets consisting of two cells (doublets), cell with insufficient data (dropouts), batch effects when analyzing samples at different times or in different ways, and many others.

When it comes to data analysis, there is also a near-infinite array of options. After obtaining the raw (FASTQ) sequencing data, most researchers will carry out a clustering analysis and examine the top DEGs. However, many other, more advanced analyses, for example functional annotations, are possible. In addition, data visualization plays a big part in communicating single cell sequencing data, and there is range of options for software well suited for visualizing single cell RNA-seq data.

All these challenges and choices can seem like a high barrier to starting up a single cell transcriptomics project. Nevertheless, the unrivalled precision and the in-depth analyses made possible with single cell sequencing mean that the efforts will almost certainly pay off as its single cell resolution truly enables new genomics discoveries.

With extensive genomics expertise and the latest sequencing technology, Macrogen Europe offers a comprehensive single cell sequencing service for researchers and industry professionals. Built on 25 years' experience in genome sequencing, Macrogen Europe's flexible offering is designed to support scientists with the most reliable data, insightful analyses and fast turnaround times.

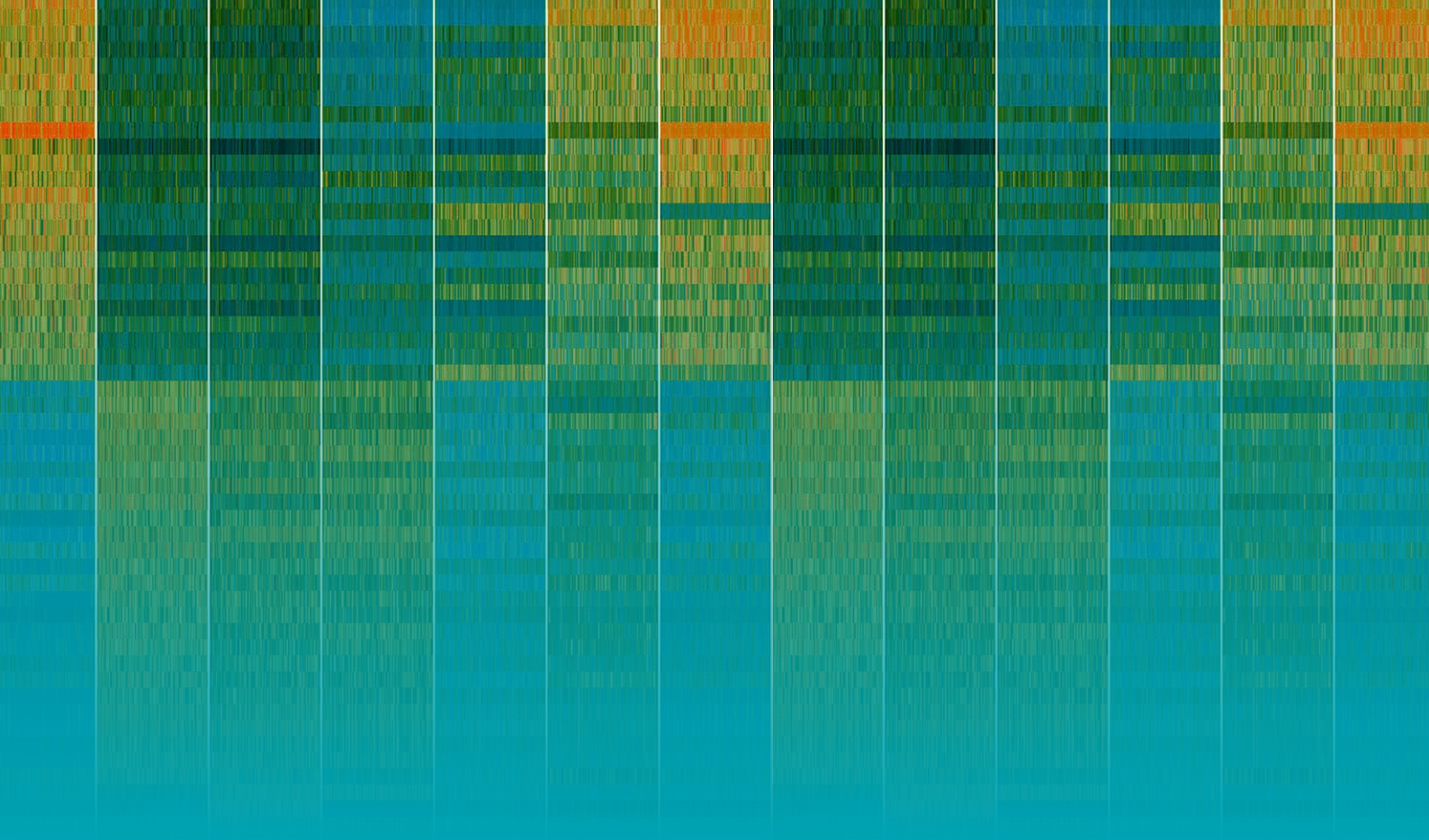
To find out more about Macrogen Europe's single cell sequencing services, visit **macrogen-europe.com** or email **ngs@macrogen-europe.com**.



References

1. Chen H and Ye F. Revolutionizing immunology with single-cell RNA sequencing. Cellular and Molecular Immunology. 2019; 16: 242–249.
2. Papalexi E and Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nature Reviews Immunology. 2018; 18: 35–45.
3. Schatz DG and Ji Y. Recombination centres and the orchestration of V(D)J recombination. Nature Reviews Immunology. 2011; 11: 251–263.
4. Neal JT, Li X et al. Organoid modeling of the tumor immune microenvironment. Cell. 2018; 175 (7): 1972–1988.
5. Scaramuzzino S, Potier D et al. Single-cell transcriptomics uncovers an instructive T-cell receptor role in adult γδ T-cell lineage commitment. The EMBO Journal. 2022; 41: e110023.
6. He H, Li Z et al. Single-cell RNA-seq reveals clonal diversity and prognostic genes of relapsed multiple myeloma. Clinical Translational Medicine. 2022; 12: e757.
7. Gibellini L, De Biasi S et al. Single-cell approaches to profile the response to immune checkpoint inhibitors. Frontiers in Immunology. 2020; 11: 490.
8. Wouters J, Kalender-Atak Z et al. Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. Nature Cell Biology. 2020; 986 (22): 986–998.
9. Ding J, Adiconis X et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nature Biotechnology. 2020; 38: 737–746.

Cover and header graphics adapted, with permission, from S Scaramuzzino, D Potier et al[4].



www.macrogen-europe.com

Macrogen Europe BV
Meibergdreef 57, 1105 BA Amsterdam
ngs@macrogen-europe.com | +31 20 333 7563

Humanizing Genomics
macrogen EUROPE