

Franziska Kasperzak

Personal information, Seminars, what I learn

Personal information: what do I expect to learn?

Hello, my name is Franziska Kasperzak, and I am currently studying Biology. I decided to take this course because I am very interested in learning more about bioinformatics, a field that is still quite new to me. I have worked with Python before, but only on a basic level, so I hope that this subject will help me understand it properly and apply it in a biological context. What I expect from this course is to gain a deeper insight into how biological data can be analyzed and interpreted using computational tools. I think it is very important to have at least a general understanding of bioinformatics, as it plays such a crucial role in modern science. Today, many areas of biology and research rely heavily on computer programs and data analysis. These tools not only save time but also reduce unnecessary manual work, making experiments and data evaluation more efficient and precise.

Since there is no course like this at my home university in Germany, I see this as a great opportunity to explore something new. I hope to get a broader overview of the different fields within biology and perhaps discover a new area that truly interests me. Even if I do not end up specializing in bioinformatics, I believe it will still be very useful to understand the basic concepts. In future collaborations, I will most likely work with people who focus on computational biology, so having a fundamental understanding of what they do will help me communicate and work with them more effectively.

Overall, I am excited to learn more about bioinformatics, improve my programming skills, and see how this field connects with the biological topics I am already familiar with.

Seminar no 1:

advances in conservation genomics driven by reference genomes

In the talk, she showed us how much conservation genomics has improved thanks to better reference genomes. With these genomes, it becomes much easier to compare individuals, populations, and even whole species. Methods first developed for model species like *Drosophila melanogaster* are now widely used in conservation.

She also told us that she never wanted to choose only one type of work. Instead, she wanted a job where she could mix field work, wet lab work, and dry lab analysis at the same time — and conservation genomics was perfect for that.

A big part of her work involves combining RNA-seq, newly collected samples, and public data from related species. RNA-seq helps them see which genes are active in the organism at a certain time. Not every species is simple to sequence, so using clear standards is important.

One example from her research was the genome of *Arbacia lixula*. This genome was created to compare it to the related *Paracentrotus lividus*. The two sea urchins differ in habitat and genetics, which helps explain how they adapt to their environments.

To collect genetic data, her team uses two main approaches:

- Whole-Genome Sequencing (WGS) for full but expensive data
- Reduced Representation Sequencing (RRS), which is cheaper and gives around 10,000 markers

Data quality is checked with BUSCO genes. Good datasets reach about 90–100%, while 50% would mean that many important genes are missing.

Another example she shared was about Mediterranean sea turtles such as *Caretta caretta*. These turtles return to the beach where they were born, and their sex depends on temperature: warmer conditions produce males, cooler ones produce females. Because the climate is changing, nesting areas are shifting, and the turtles are moving into cooler regions.

At the end, someone asked about her favourite species. She explained that she doesn't have one, because every species is interesting and has something special to discover.

Seminar no 2:

Introduction to Single-Cell Technologies

Irepan Salvador is a researcher at the Barcelona Supercomputing Center (BSC) in the Transcriptomics and Functional Genomics Lab led by Marta Melé. His work focuses on modern transcriptomic approaches, particularly single-cell technologies, which have transformed the way we understand cellular diversity and tissue function.

In his presentation he talks about different technologies. One of them was the bulk RNA-seq. Bulk RNA sequencing was developed in the early 2000s as one of the first high-throughput transcriptomic methods. In bulk RNA-seq, RNA from thousands or millions of cells is extracted together, producing an averaged expression profile. This approach is cost-efficient and widely used for comparative transcriptomics, biomarker discovery in disease studies, and the analysis of homogeneous systems. However, because all cell types are mixed, bulk RNA-seq cannot resolve cell-type-specific gene expression.

The other technology was single-Cell RNA-seq (scRNA-seq).

Irepan Salvador explained, that Single-cell RNA sequencing emerged around 2009 and enabled researchers to profile gene expression at the resolution of individual cells. Technologies such as the 10x Genomics Chromium system introduced GEM (Gel Bead-in-Emulsion) microfluidics, allowing high-throughput partitioning of single cells and construction of individual sequencing libraries.

Irepan Salvador then introduced Single-Cell RNA Sequencing, commonly referred to as scRNA-seq, as one of the most important modern transcriptomics technologies. He explained that this approach emerged around 2009 and marked a major breakthrough compared to traditional bulk RNA sequencing. For the first time, researchers were able to measure gene expression at the level of individual cells, rather than averaging signals across large populations of cells.

As a key example, Irepan highlighted the 10x Genomics Chromium system, which uses GEM (Gel Bead-in-Emulsion) microfluidic technology. In this system, single cells are encapsulated together with uniquely barcoded gel beads inside tiny droplets. This enables the construction of separate sequencing libraries for thousands of individual cells in parallel, making high-throughput single-cell analysis possible.

He emphasized that the fundamental conceptual difference between bulk RNA-seq and scRNA-seq is that, in single-cell experiments, each library corresponds to a single cell, rather than a mixture of many cells. This allows researchers to distinguish different cell types, visualize them as distinct “colors” in dimensionality reduction plots, identify rare cell populations, and explore cellular heterogeneity with unprecedented resolution. However, Irepan also pointed out that despite its power, single-cell sequencing is considerably more expensive than bulk RNA-seq approaches.

Irepan then outlined the key analysis steps involved in single-cell RNA-seq data processing. These include quality control of cells and genes, normalization of gene expression, selection of highly variable genes, dimensionality reduction using methods such as PCA, UMAP, or t-SNE, clustering of cells, cell-type annotation, and differential expression analysis. He also mentioned optional downstream analyses, such as trajectory or pseudotime analysis and cell-cell interaction analysis.

Finally, Irepan moved on to Spatial Transcriptomics. He explained that while scRNA-seq reveals which cell types are present in a sample, it loses information about where those cells are located within the tissue. Spatial transcriptomics addresses this limitation by preserving the spatial coordinates of RNA molecules during sequencing or imaging. This technology measures gene expression in intact tissue sections, assigns spatial barcodes to mRNA molecules, and allows researchers to reconstruct two-dimensional maps of gene expression across tissues. According to Irepan, spatial transcriptomics enables the study of tissue architecture, cellular microenvironments, and cell-cell interactions *in situ*, effectively combining the strengths of histology and single-cell sequencing. He mentioned

technologies such as 10x Genomics Visium, Xenium, MERFISH, and related platforms as prominent examples in this rapidly evolving field.

He used a simple LEGO-based illustration to explain the differences between transcriptomic technologies:

Bulk RNA-seq:

All LEGO pieces of different shapes and colors are mixed together in one large pile.

Single-Cell RNA-seq:

The pieces are separated into multiple smaller piles, sorted by color—each pile representing a distinct cell type.

Spatial Transcriptomics:

The LEGO pieces are arranged in a flat 2D pattern, preserving their positions, like a map of the tissue.

Functional Tissue (3D organization):

The LEGO bricks form a fully assembled 3D structure, showing how cells are arranged to create a functional organ.

Irepan Salvador told us about the Xenium (10x Genomics).

Xenium is a high-resolution *in situ* transcriptomics platform. He explained how it works:

1. Padlock probes bind to the target RNA sequence within the tissue.
2. The probe is circularized and a copy of the target sequence is synthesized by rolling-circle amplification.
3. Fluorescent barcodes are repeatedly hybridized and imaged.
4. Thousands of RNA molecules are detected directly in their spatial context with subcellular precision.

One question that has stuck with me is how the salary would be. His answer shocked me and honestly put me off at first. The salary was very low. I thought about it again and realized that you can't directly compare German and Spanish salaries. Even though it's not just about the salary in the end, I still think it's important to pay attention to it. Irepan Salvador said that too, because it's important to have enough pension in the end to start a family or whatever.

What did I learn in this seminar?

In this module, I learned many new skills. First, I improved my Python coding. I didn't just learn the commands, I learned how to actually use and apply them to solve given problems. A big part of the class was reading scientific papers. I learned how to read these difficult texts, understand what they want me to do, and then use the new knowledge.

I also learned talk if you struggle with anything. Instead of giving up, we asked each other and solved problems as a team. This made the work easier and helps everyone. Now I feel much more confident in my coding.