#RESEARCHNEVERSTOPS

Evotec

… perspective for Computational Biology and Bioinformatics talents in Drug Discovery

Evotec SE, Perspectives in Biology lecture series at University Göttingen, March 2022

Evotec's Data-Driven R&D Autobahn to Cures

Transforming the way how drugs are discovered and developed

Fully integrated

Target ID/ validation → Hit identification → Lead optimisation → Pre-clinical development → Clinical & commercial manufacturing

All-modality design

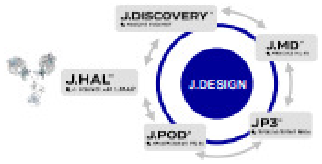
EVO*cells* / **EVO***genes* / Antibodies & Bifunctionals
Antisense / RNA / Exosomes / Protein degradation
Small molecules

Higher efficiency

(30%) time

(50%) costs

AI enabled disruptive biologics



iPSC & AI for higher PoS

Co-owning strategy

>150

Partnered assets
EVO*equity participations*
BRIDGEs

Omics-driven precision medicine

EVO*panHunter*

EVO*panOmics*

Fast growing and highly profitable

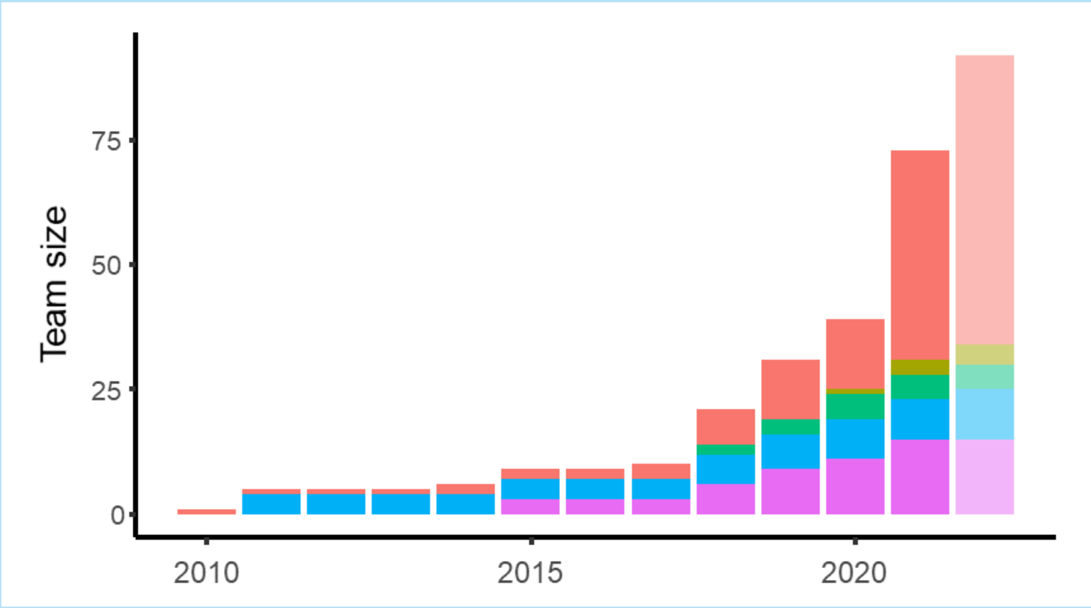
> € 500 m
Revenues 2020

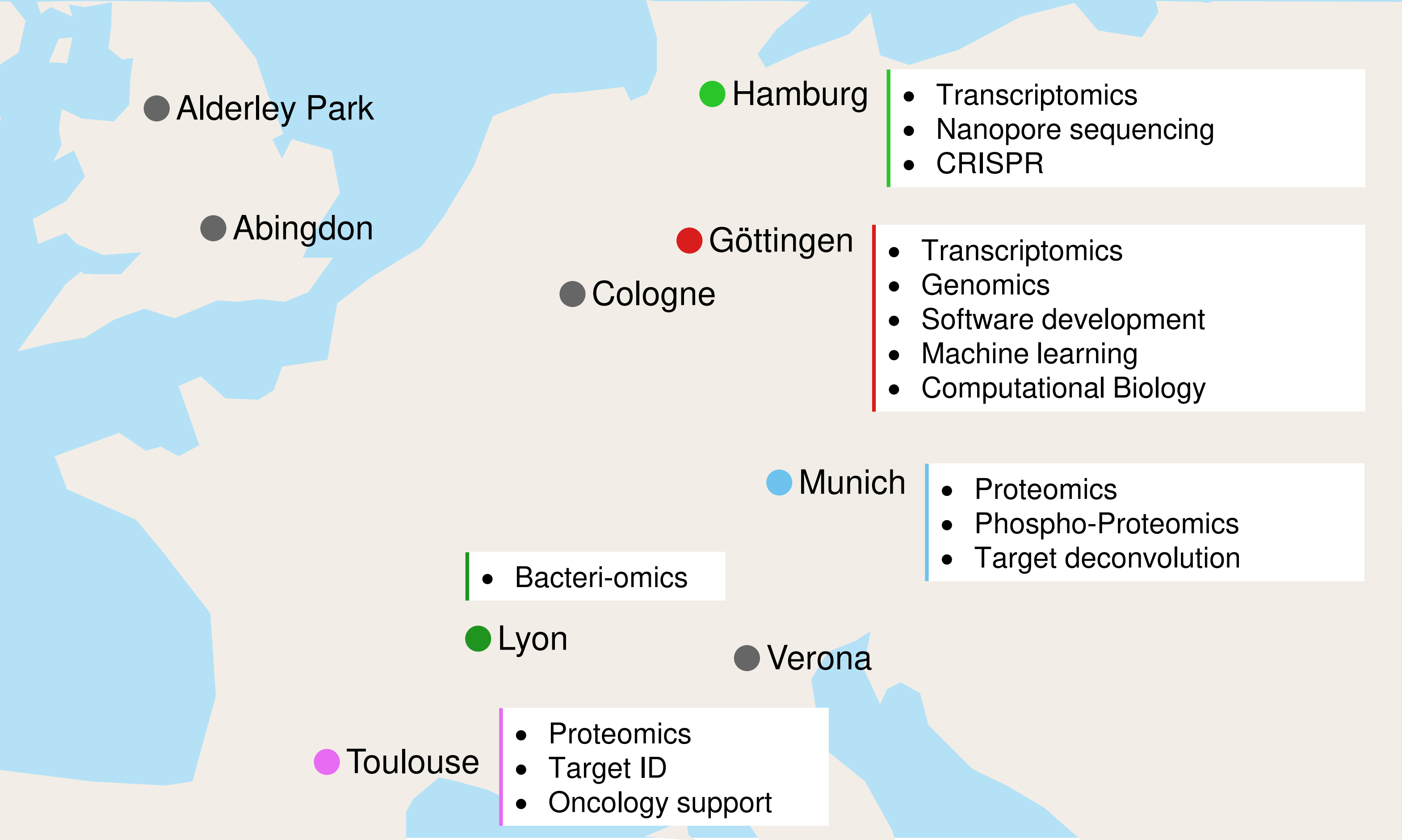
> € 100 m
Adj. EBITDA 2020

Revenue CAGR goal of > 15%

BIX at Evotec: strong growth of *in silico* teams along with overall company growth

Diverse backgrounds in biology, medicine, computer science, statistics, and math





Hamburg

- Transcriptomics
- Nanopore sequencing
- CRISPR

Göttingen

- Transcriptomics
- Genomics
- Software development
- Machine learning
- Computational Biology

Munich

- Proteomics
- Phospho-Proteomics
- Target deconvolution

Lyon

- Bacteri-omics

Toulouse

- Proteomics
- Target ID
- Oncology support

A diverse team of data analysis experts delivers insights into complex omics data

From raw data to the final figure

Planning phase

- Support of experiment design

Method development

- Project-specific applications

Data analysis

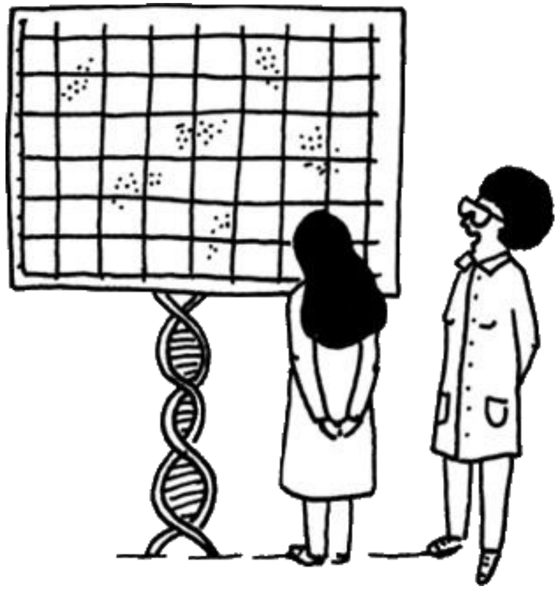
- State-of-the-art methodology

Data interpretation

- Disease & Biology experts

Scientific reports

- Publication-ready insights



Data analysis & interpretation

Educational Background

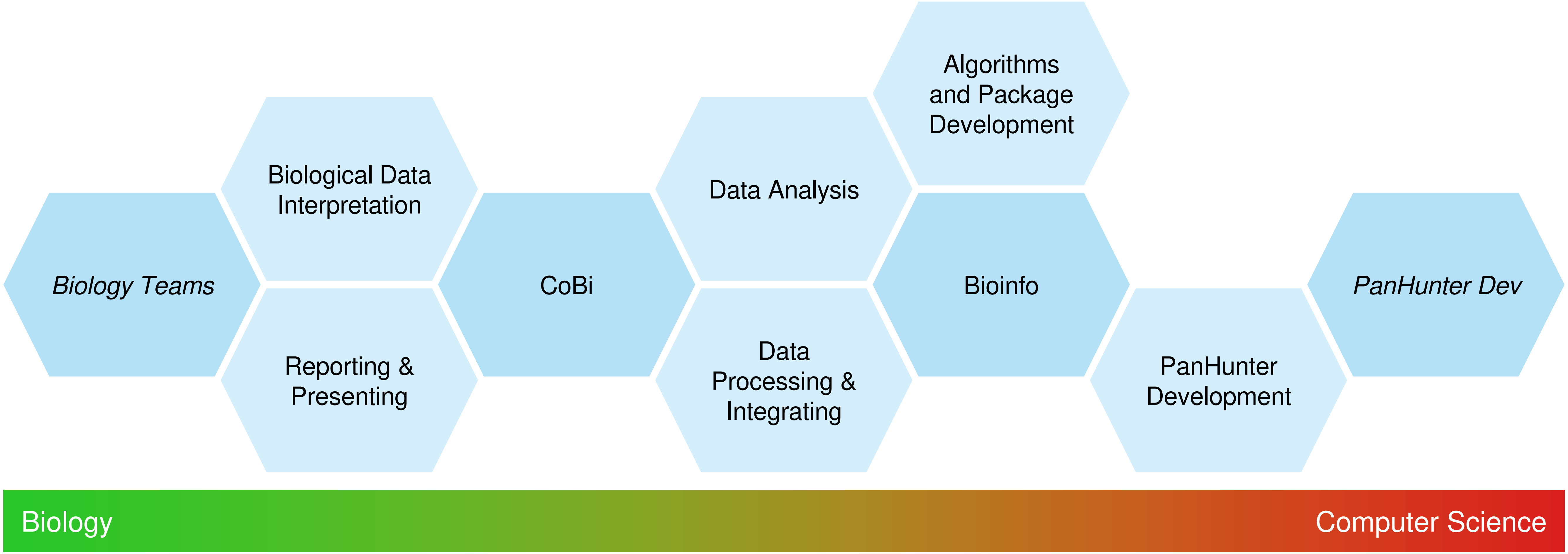
- Computer Science
- Mathematics
- Biochemistry
- Bioinformatics
- Biology
- Biomedicine
- Pharmacology

Scientific profiles

- Genomics
- Metabolomics
- Programming
- Proteomics
- Single-Cell and Spatial Technologies
- Software development
- System engineering
- Transcriptomics
- Aging
- Cardiology
- Developmental Biology
- Immunology
- Infectious diseases
- Metabolic diseases
- Neuroscience
- Oncology

Task-Diversity and interplay between Evotecs BIX teams

Integrated, cross-functional teams covering all aspects from Biology to Computer Science


Biology Teams

Biological Data Interpretation

Reporting & Presenting

CoBi

Data Analysis

Data Processing & Integrating

Algorithms and Package Development

Bioinfo

PanHunter Development

PanHunter Dev

Biology

Computer Science

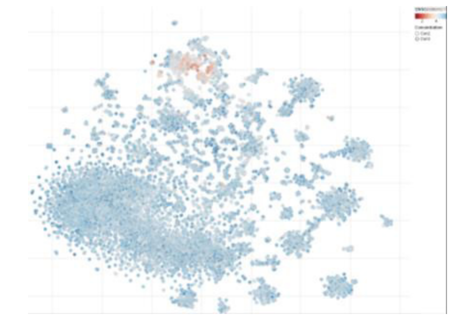
What data do we work with and what scientific question do we try to answer?

Transcriptomics *in vitro*, *in vivo* and *in sito* is our main technology in Göttingen

- **Bulk RNA-Seq on Patient Cohorts**
  - Profiling the blood and tissue Transcriptomes of patient cohorts (1-10k patients, longitudinal follow-ups) with a disease
  - Often complemented with selected, other omics data like genomics or proteomics – either for full cohort or selected part of the cohort
- **Disease Understanding**
  - What are the aberrations that differentiate a disease patient from a healthy patient?
  - What do these aberrations tell about the disease? What are the underlying deficiancies?
- **Patient Stratification**
  - What groups of patients do we have in our cohorts?
  - What are the similarities and dissimilarities between the group of responder/non-responder or disease-progression and stable-disease?
- **Biomarker ID**
  - Can we find a biomarker (individual omics feature) or a signature (a combination of omics features) that reveal the state of the disease progression?
  - Can we find a biomarker (individual omics feature) or a signature (a combination of omics features) that reveal the expected outcome, e.g., faster decline of kidney function
- **Target ID**
  - Can we identify proteins that are aberrant to a control or in later stages and build a mechanistic understanding how this protein could drive the disease progression?
  - Could this protein be targeted by a drug and how?

- **Transcriptomics Screening using Evotec ScreenSeq™**
  - In-house high-throughput RNA-Seq protocol ScreenSeq™ enables the generation of thousands of samples (highly multiplexed and shallow sequenced)
  - Primary use in investigating compound effects in cell lines, at different concentrations, at different time points, etc.
  - Data sets comprises samples for thousands of compounds, easily summing up to 25k or 50k individual samples
- **Mode of Action**
  - How does a specific (group of) compounds work in this cellular model?
- **Disease Reversion and Mitigation**
  - Which compounds revert the disease phenotype (induced) most efficiently?
- **Toxicity**
  - Which of the compounds show effects that are associated with toxicity or other safety related aspects? An early de-prioritization saves cost later in the drug discovery process.

Transcriptomics Screen with 23k samples



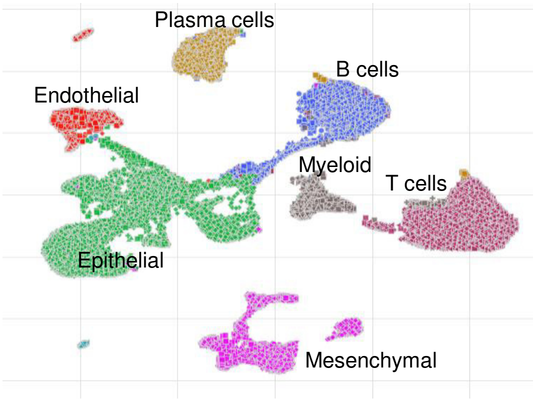
What data do we work with and what scientific question do we try to answer? (cont’)

Platform technologies that enable the discovery of new modalities and treatment options

Single-cell, single-nuclei RNA-Seq & Spatial Transcriptomics
Various types of high-resolution Transcriptomics are routinely used across the various research areas. Examples include:

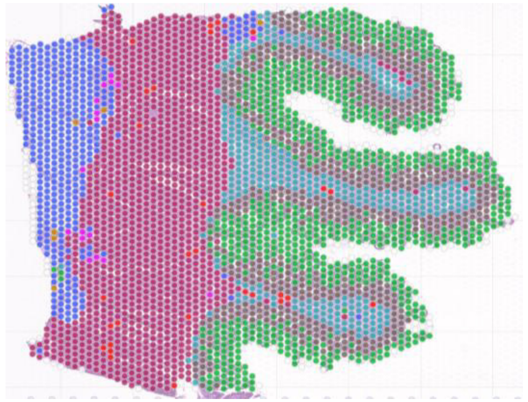
- Diseases
  - Improve our understanding of the underlying diseases we develop drugs against
- iPSCs
  - Track the differentiation process when developing disease models
  - Investigate compound effects on the single-cell level

sxRNA-Seq with ~10k cells



Spatial context

Spatial Transcriptomics
(5k spots each with ~10 cells)



Whole-Exome and Whole-Genome Sequencing
DNA sequencing data is primarily used for understanding the structural changes of our cellular models as in iPSC, cell model, and cell-therapy QC:

- Which off- and on-target effects resulted from our CRISPR editing experiment?
- Does our cell line maintain genetic stability of passaging or does it develop mutations which give a proliferation advantage (carcinogenic)?

But large-scale efforts like UK BioBank demonstrated the value of sequencing over genotyping and we start to explore these technologies for large-scale cohort analyses too.

Cohort Genotyping and GWAS
Genotyping with SNP arrays enables a cost-efficient way to screen the genetic markup of large populations

- Patient & Humanomics
  - What SNPs are prevalent in disease under investigation?
  - Can we stratify the patients-cohort e.g., by pre-disposition?
- iPSC & cell model QC:
  - Routine use for cost-efficient screening for copy-number variations

What data do we work with and what scientific question do we try to answer? (cont')

Core Bioinformatics work on packages, algorithms, and methods for data analysis

How can we get the most out of our data?

- Develop **data processing pipelines** for data types not yet available at Evotec. This includes evaluation and benchmarking of existing tools. Where tools are missing or not fast enough, we develop own tools.
- We develop best practices, quality control and our internal gold standards
- Future topics will include
  - Methylation
  - Metabolomics
  - Metagenomics
- We constantly **improve existing pipelines** with faster or more accurate methods:
  - gene expression quantification
  - transcript expression over gene expression
  - spatial technologies
  - variant calling
  - genotyping
  - etc.

How do we make sense out of our data?

- We develop the scientific core of the **EVOpanHunter data analysis** platform and extend it with our analysis tools & packages
  - Fast **Exploratory Analysis** tool enables screening clusters/groups of samples against associated meta-data and features (gene expression). It identifies the drivers of complex dimension reductions.
  - **evoEnrichment** analysis properly implements gene set enrichment on ontologies (e.g., Gene Ontology) and nested groups (e.g., Reactome pathways) at scale. Enables the executing of pathway enrichment on thousands of comparisons (like in ScreenSeq data sets) in proper time.
  - The **Community Analysis** enables the mining of thousands of comparisons (e.g., different compound effects) to identify communities of genes/proteins with similar co-regulation patterns
  - Our **single-cell-classifier** allows the annotation of millions of cells by reference samples. Transferring of cell type and disease states across data sets improves the time to subsequent analysis for biology experts.

Check out the "EVOpanHunter" playlist on YouTube

Frequently Asked Questions – Getting the job

Make sure you have the required hard-skills and put effort in your application

What hard- and soft-skills are essential for your work? What additional requirements are relevant for positions in your team (explicit skills, internships, working abroad, …)? What prior knowledge is relevant?

- *Prior experience in working with OMICs data is important (ideally larger >100 samples). We have a focus on Transcriptomics, but with an increasing number of Proteomics, Genomics, and Metabolomics data sets coming in.*
- *We hire on PhD and Master-level!*
- ***Hard-skills for Bioinformatics***
  - *good understanding of programming (ideally R or Python)*
  - *some experience in R package development or general software-development is highly appreciated*
- ***Hard-skills for Computational Biology***
  - *You are passionate about solving complex puzzles and have a strong understanding of metabolic diseases, inflammation, or pharmacology*
  - *You have experience with analysing high-dimensional, biological data*
  - *You don't need to program, but you should be trained in applying computational tools, e.g. R packages, to mine data*
- *Required **soft-skills** are a good self-management (tasks & time), good communication skills with all levels of staff (peers to management), good presentation skills, i.e. ability to deliver the keys messages from complex analysis to stakeholders and non-experts*
- *You like*
  - *working on diverse and international teams; Fluency in English is required*
  - *developing your skills: Evotec aims to close the gap between state-of-the art basic/academic research and industrial application – and research never stops!*

How did you manage to get hired in industry? What was particularly important?

- We (and others) screen CVs for **relevant hard-skills that are also clearly communicated in the job advertisements**. Highlight your experience in those – don't let me search for it!
- Do not underestimate the impact of your CV and CL appearance to the recruiter and hiring manager. A **personalized CL & CV** demonstrates that you read the job-ad. A personalized sentence about **your motivation** on applying to this position shows that you thought thoroughly about the job & the company.
- During the personal interviews, you will also be strongly evaluated by your soft-skills and by your personal attitude. It is highly important that your personality matches that of the team because otherwise it will generate immense friction. This is a two-way street – if your future colleagues do not share your ideals, this will be a tough job for you!

Frequently Asked Questions – the future

Evotec is a fast-paced research organization with great opportunities for growth

How does a typical working day look like?

- *All work as CoBio or Bioinfo is done in silico. We work closely with the wet lab colleagues, but no need to touch a pipette :-)*
- *On average we have 2-4h meetings per day to align within the teams, the projects or globally. The rest of the day is left with project work.*
- *For Bioinfo, this entails*
  - *1/5 data processing*
  - *1/5 planning & concept discussions*
  - *3/5 programming for data analysis (scripts), scientific packages, pipelines*
- *For Computational biology it is all about solving puzzles, you spend*
  - *1/3 of your time analysing data*
  - *1/3 of your time brainstorming with other team members*
  - *1/3 of your time presenting data to projects or clients*

How is a career path?

Evotec uses a Matrix Organisation and offers two different career paths:

1. *Line Management – responsibility for teams with specific functions:*
   - *Your job is to develop the scientific core of your team,*
   - *organize work packages, resources,*
   - *and – most importantly – develop your team members. Similar to a core facility.*
2. *Project Management – responsibility for (part of) a drug discovery project:*
   - *You develop and guide internal RnD projects, or*
   - *manage external "fee for service" projects with cross-functional project team.*

How is the work-life balance?

- *Evotec Germany employs with 40h per week, based on the trust principle, i.e. working time is not tracked. Overtime, however, is tracked. You are responsible to manage your time. Mobile work is possible.*
- *If you are able to manage your tasks well, a good work-life balance is no problem. After your work, you shut down the computer and there is also no expectation that you reply to emails late. If you need to stay longer, you compensates with “Out of office”-time the next day. If we have a private appointment, we simply put out-of-office in our calendar, reschedule meetings and take the private appointment.*

How much money to earn?

Evotec salary ranges are between TV-L and "Big Pharma". The salary consists of a fixed annual salary plus several performance-based components.

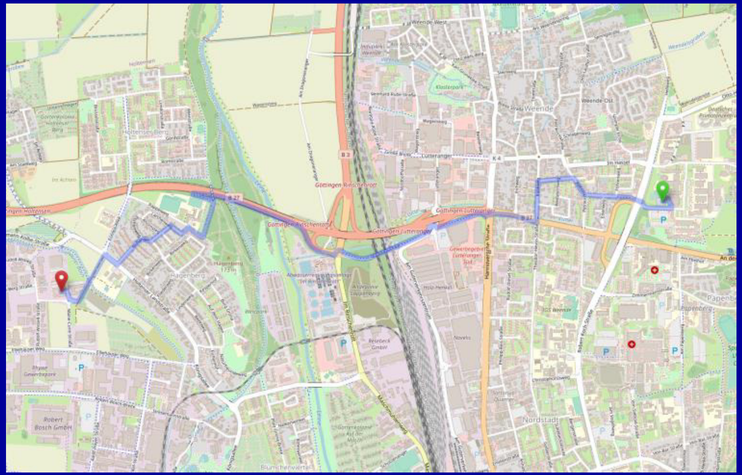
Questions after the talk

- **Does Evotec offer opportunities for Master- or PhD-Thesis?**
  - *Not done regularly and many projects are not suited because projects are associated with external partners imposing strict time-lines and sometimes also shifts in work-packages. Opportunities rather exist within our internal R&D projects. If you have a topic that you believe fits our working area and you like to do a industry-backed thesis, please contact us and we are happy to discuss (see last slide).*
- **Does Evotec offer opportunities for Internships?**
  - *Also here, we do not have this regularly but we are open to do. To really understand more of the ways we work and benefit from your internship, we recommend internships of about 3 months. If you are interested to join for longer, we can also explore working-student contracting. Again – please contact us to discuss in depth ☺*
- **Is formal education through PhD and Post-Doc necessary for your positions?**
  - *Not necessarily. But to be honest, it is easier to get a job in the overall drug discovery industry if you have a PhD. But in my opinion it is also a personal choice if you want to make your way through PhD – and it is fine if you don't want that ☺*
  - *We hire Master-level graduates. So far primarily as "Scientific Programmers" focusing more on the technical aspects in the core-bioinfo teams. From a career-perspective, with the training and experience you will receive at Evotec, you will end up in a similar level as a entry-level PhD in similar overall time (3-4y PhD versus 3-5y Scientific Programmer at Evotec → same grade; that is a rough estimate and clearly depends on individual persons interest and performance). There is no "hard-career-stop without a PhD" (at Evotec; don't know about other companies).*
  - *With a Master entry you will have a more supportive role in the projects whereas with a PhD entry, you are expected to be able to manage yourself and organize your part of the project-work completely on your own. Also for PhD we expect much more experience with reporting and presenting scientific results to diverse audiences.*

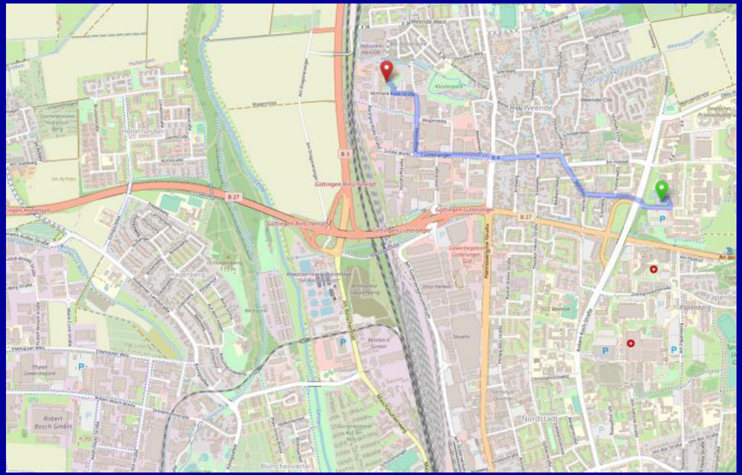
evotec

#RESEARCHNEVERSTOPS

Evotec – Main Site in Göttingen



Evotec – Klosterpark (from April 1st)



Your contact:

Manuel Landesfeind (Bioinformatics)

manuel.landesfeind@evotec.com

Sven Sauer (Computational Biology)

sven.sauer@evotec.com