Biographica

Accelerating crop improvement with AI

A platform for accurate & efficient target discovery
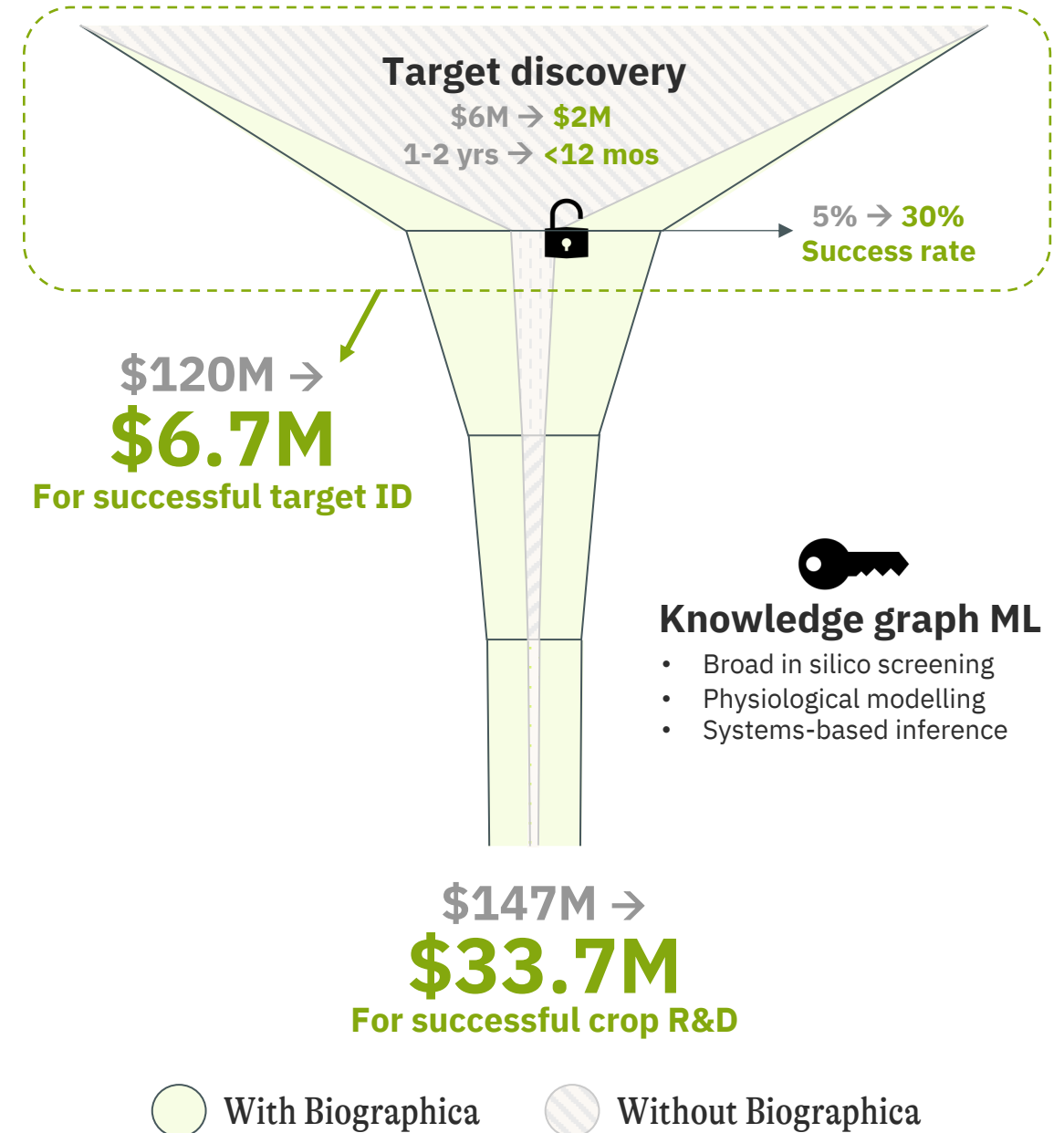
March 2024

What we do

Reshaping the crop development funnel

Target discovery is a major bottleneck in gene-edited crop R&D. Our mission is to streamline editing pipelines using machine learning (ML)

- ✓ **6x** higher target discovery success rates
- ✓ **2-3x** faster target discovery cycles
- ✓ **18x** lower costs for discovery of a successful target

1. Bullock et al. (2021) Amer. J. Agr. Econ.

R&D pipeline for gene-edited crops[1]

Target discovery
$6M → **$2M**
1-2 yrs → **<12 mos**

5% → **30%**
Success rate

$120M →
$6.7M
For successful target ID

Knowledge graph ML
- Broad in silico screening
- Physiological modelling
- Systems-based inference

$147M →
$33.7M
For successful crop R&D

With Biographica   Without Biographica

2

How we do it

Our discovery platform: decoding crop genetics with AI

Current capabilities
Gene ID

Next steps
Perturbation ID

Mature platform
Holistic editing strategy

2024-2025

2025-2026
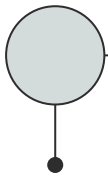
Ranked list of lead genes for given crop trait in weeks
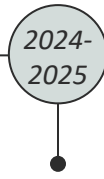
Gene perturbations for **quantified** trait effects
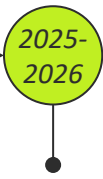
Edit / allele combinations for quantified trait effects

"Gene x then gene y are top-priority targets for trait z"

"Down-regulate gene x for y% increase in trait z"

"Combine edits a & b for y% increase in trait z"

Current capabilities

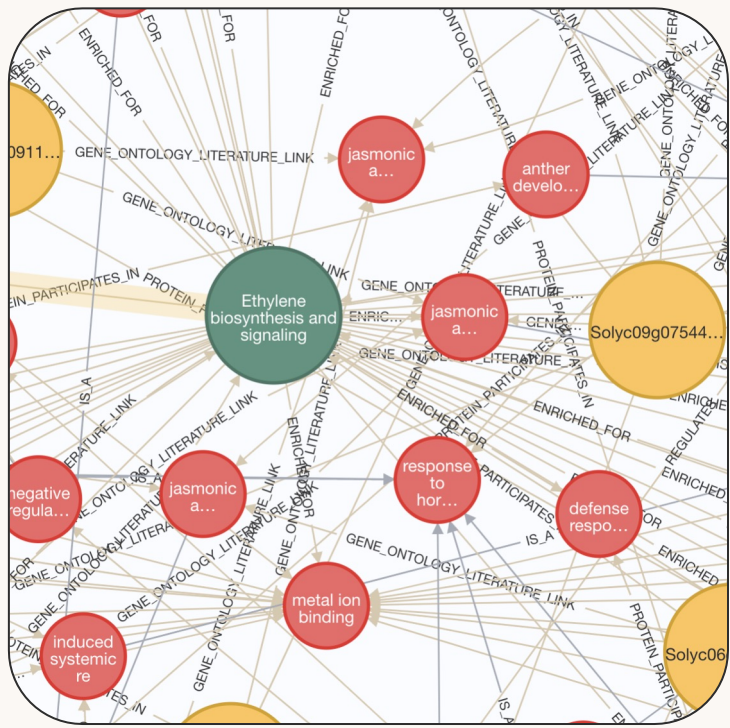An iterative platform for accurate, rapid & cost-effective gene identification

1 Data integration
- AI-driven literature search
- AI-driven functional annotation
- Automated integration of public data
- Dynamic integration of experimental data

2 Graph construction
Systems-based modelling of gene-trait networks

3 Target identification
Computational prediction of gene targets

Crop & trait of interest

Novel, prioritised targets in ~2wks

Customised to customers' lines

Pleiotropy & biological pathway analysis

Discovery timeline
(Days)

1   14

*Currently developing pipeline to predict perturbation strategy for desired phenotypic effect (e.g. upregulate gene x for increase in trait y); Q1 2024 forecast completion date

Successful proof-of-concept for gene ID in tomato

Custom graph construction*

Project deliverables

Report:
- Ranked target list
- Pleiotropic analysis
- Supporting evidence

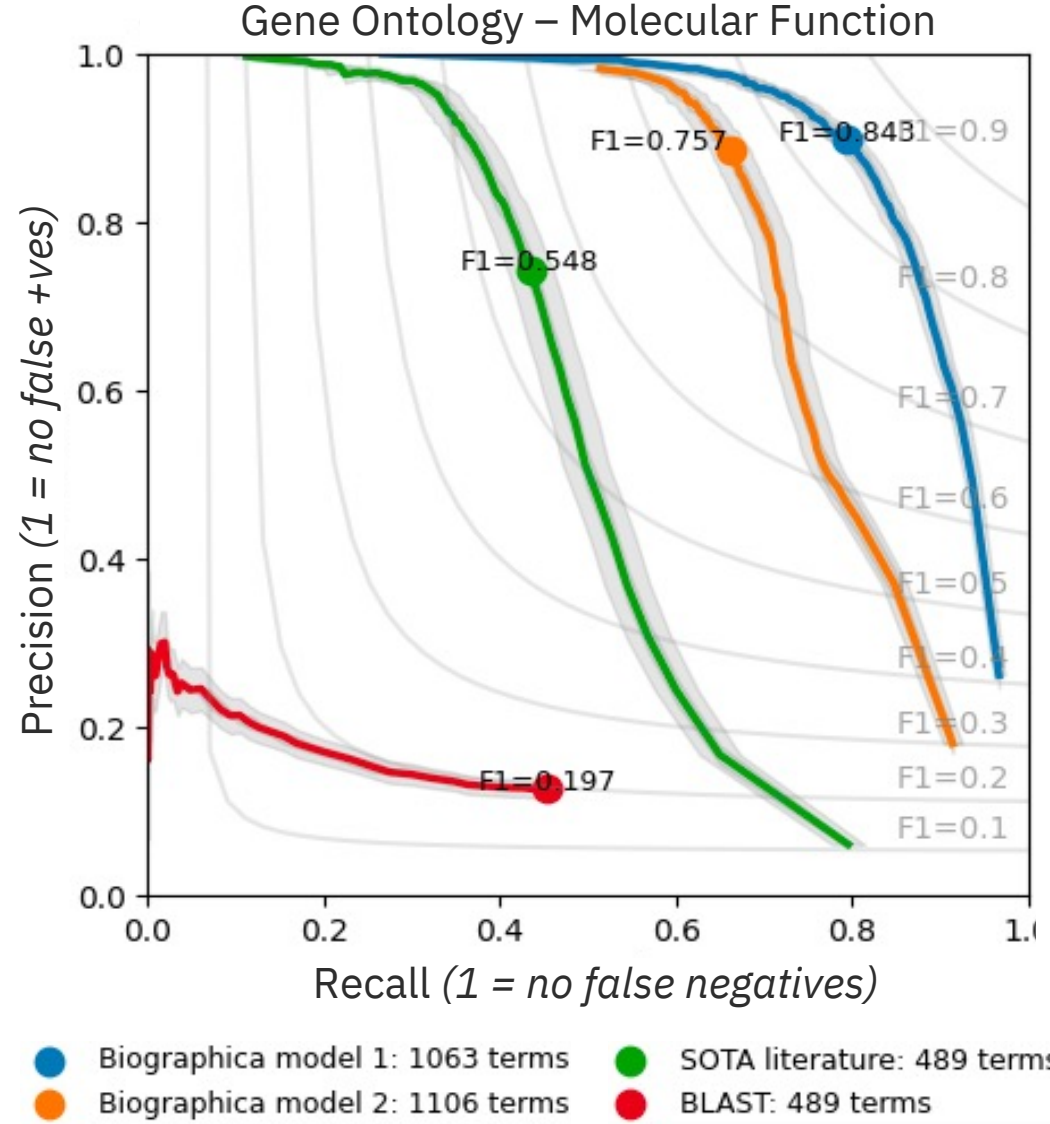Subgraphs to support hypothesis generation & explainability

Project outcomes

- Recalled **7/7 'ground truth' targets** for trait

- Identified 14 further high priority targets

- Targets to be validated experimentally starting Q4 2023

State-of-the-art Protein Function Prediction

...DNVGK
ADDHRT...

Protein sequence

Biographica →

GO annotations

Vs other published methods, we predict:
- **More annotations** per plant protein
- With **higher accuracy**
- **From protein sequence** alone

Highly accurate, proprietary annotations:
- **Boost success rate** of target ID
- **Reveal biological mechanisms** of targets

We outperform BLAST by 328% at prediction of plant molecular functions

Gene Ontology – Molecular Function

F1=0.843

F1=0.757

F1=0.548

F1=0.197

Precision (1 = no false +ves)

Recall (1 = no false negatives)

- ● Biographica model 1: 1063 terms
- ● Biographica model 2: 1106 terms
- ● SOTA literature: 489 terms
- ● BLAST: 489 terms

F1max for classification of molecular function GO terms; a score in the top right corner would represent a perfect model which classifies all GO terms correctly for a given protein

Validated Discovery of Drought Resistance Target
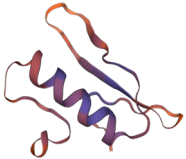
Our protein function prediction models confidently predict novel* gene-trait links **(e.g. ABR1** *see right)*

Implications**:**

1. Predictions confidently diverge from public databases → **1000's of novel potential targets**

2. We only use sequence → **can annotate arbitrary protein variants in your germplasm**

3. Annotation is just one tool → **also tell you about efficacy, pleiotropy & evidence for each target**

Biographica Predicts Novel Function for ABR1 in Drought Response

ABR1	GO:0003700 DNA-binding TF activity	GO:0009414 Response to water deprivation
pBLAST	✅	**??**
Uniprot (A0A178UFQ7)	✅	**??**
	✅ 99% confident	✅ 95% confident

2023 Publications Validate Biographica's Discovery

- *PtrABR1* Increases Tolerance to Drought Stress by Enhancing Lateral Root Formation in *Populus trichocarpa*

- Transcription factors ABF4 and ABR1 synergistically regulate amylase-mediated starch catabolism in drought tolerance

Where we're at

Milestones to date

Commercial POC
First commercial contract for POC with UK seed company
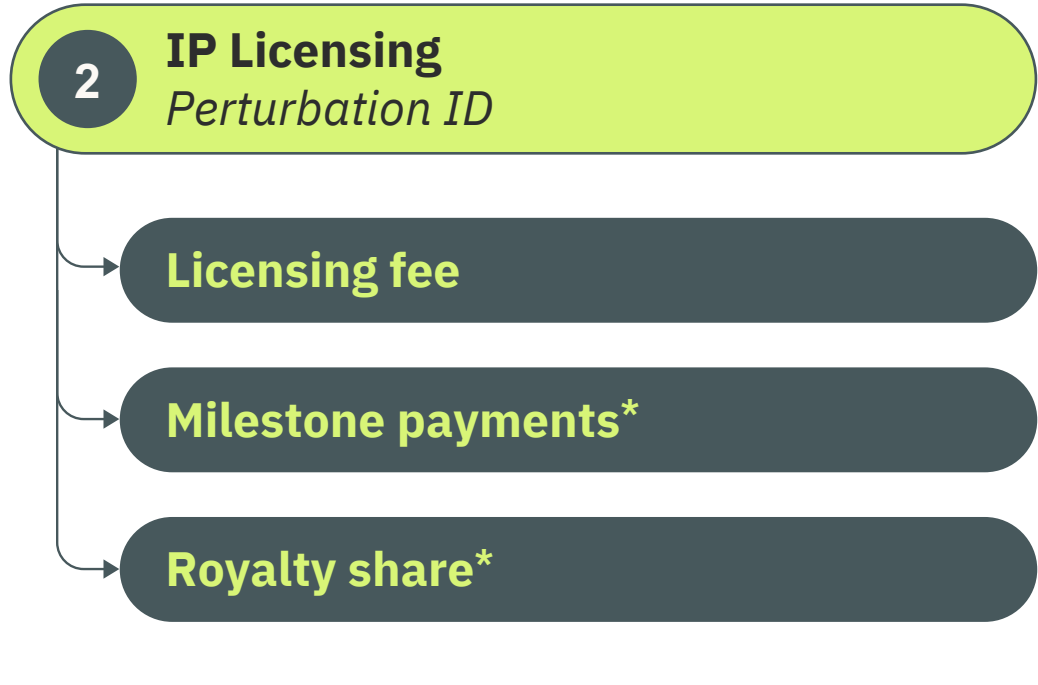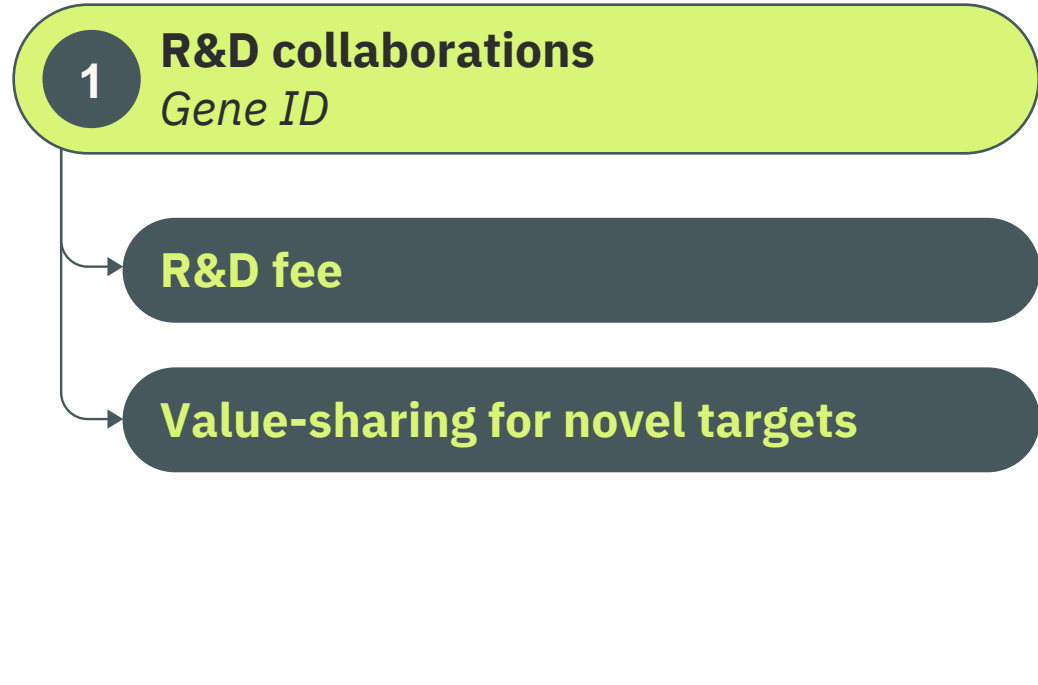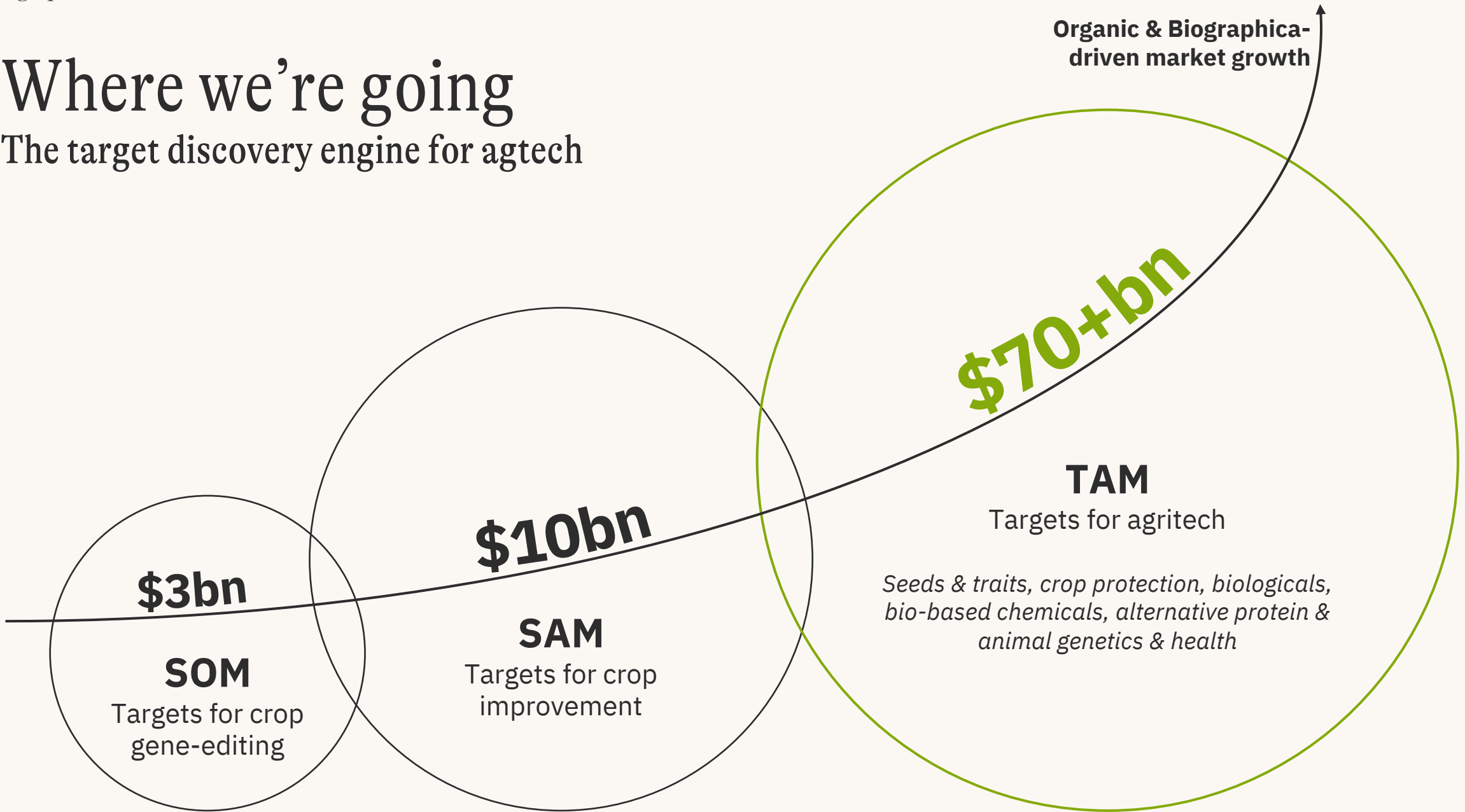
MVP Complete
Ability to ID ranked gene lists for any crop trait in ~1mo

Syngenta Shoots
1 of 7 startups selected to join Syngenta's Shoots Accelerator

Pre-seed Raise
Close £1.6M pre-seed round led by Chalfen Ventures

POC Success
Recall 7/7 "ground truth" targets for commercial partner in POC

Incorporate

Nov '22 → Mar '23 → May '23 → Jun '23 → Jul '23 → Aug '23

Apr '24 ← Mar '24 ← Dec '23 ← Nov '23 ← Oct '23 ← Sep '23

2 New Scientists
Oxford postdoc in disease resistance gene ID; ex-BASF & -KWS gene ID

Commercial Pilots
2 commercial partnerships with leading seed & trait companies

Novel Target ID
ID novel drought resistance target from protein sequence alone

Efficiency Gains
Gene ID in 1wk & customisable to genetic contexts of customers

Accuracy Gains
SOTA models for protein function prediction boost gene ID accuracy

3 New Engineers
2 ML PhDs (Oxford & UCL); >10yrs target discovery experience

Business model

Development & out-licensing of target IP

1 **R&D collaborations**
Gene ID

- R&D fee
- Value-sharing for novel targets

2 **IP Licensing**
Perturbation ID

- Licensing fee
- Milestone payments*
- Royalty share*

Flexibility on relative structure of value-sharing components

Where we're going
The target discovery engine for agtech

Organic & Biographica-driven market growth

$70+bn

$10bn

$3bn

SOM
Targets for crop gene-editing

SAM
Targets for crop improvement

TAM
Targets for agritech

Seeds & traits, crop protection, biologicals, bio-based chemicals, alternative protein & animal genetics & health

Who we are

Combining expertise in plant sciences, target discovery & machine learning

Cecily Price, *Co-founder & CEO*
- Oxford Biological Sciences specialising in Genetics
- Drug manufacturing at GSK, market access strategy consulting for gene-editing technologies
- Imperial MSc, ML & AI specialising in graph ML for biological modelling

Dominic Hall PhD, *Co-founder & CTO*
- Cambridge Masters, Maths, Stem cell Biology
- Cambridge PhD, Computational Genomics focusing on graph ML for gene regulation prediction
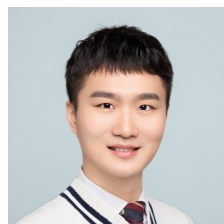- Graph ML research for gene discovery in cancer cell lines at Relation Therapeutics

---
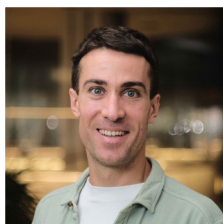
Paride Antinucci PhD
ML Engineer

Molecular Biology PhD (King's College London); BenevolentAI graph ML for target ID

Guy Aglionby PhD
ML Engineer

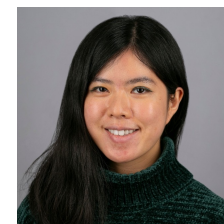PhD in Natural Language Processing using graph ML (Cambridge)

Linhao Luo
ML Researcher

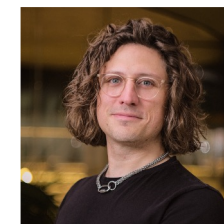Computer Science PhD (Monash) focussed on LLMs, knowledge graphs & graph ML

Felix Homma PhD
Plant Scientist

Plant disease resistance gene ID - PhD, Postdoc (Oxford) & industry (BASF & KWS)

Victoria Auyeung
Computational Biologist

Metabolomics PhD (Cambridge); plant disease resistance research (Sainsbury Lab)

Dennis Schwartz
Data Engineer

Bioinformatics MSc (TU Munich); 7+ years data engineering for graph-based target ID

Biographica

Cecy Price, Cofounder & CEO
cecy@graphica.bio | +44 7943 298499 | www.graphica.bio