

Ara-BOX-cis v2:
A **G-box** regulatory **network** for
a **single cell developmental atlas**

Daphne Ezer

Check-in code: 227956

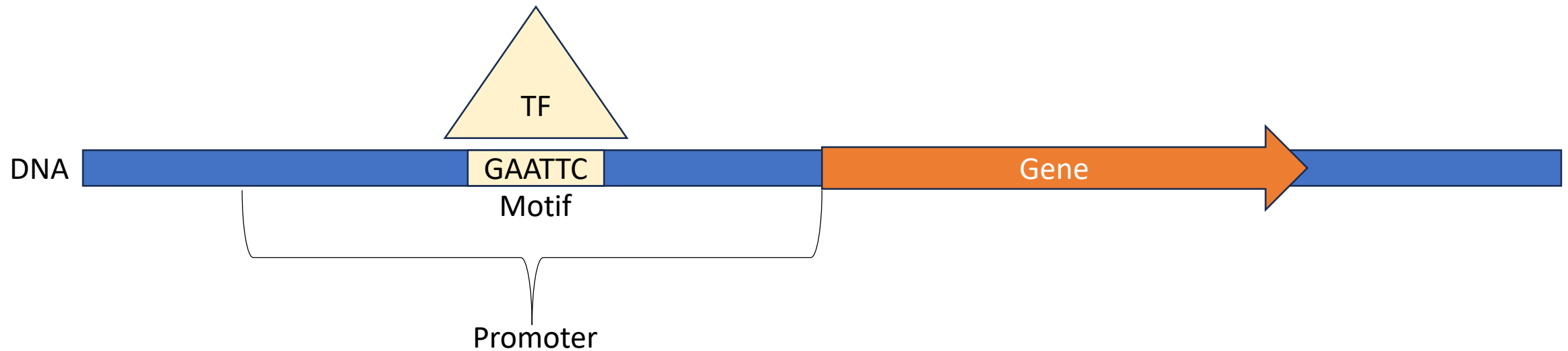
Outline of this term's material

1. Overview of project: what is the main aim?
2. Background on **G-box** regulatory networks
 1. What is a gene network?
 2. What is machine learning?
3. Background on the **data set**
 1. What is special about single cell RNA-seq?
 2. What are some exploratory analyses and visualisations we can do?
4. Background on **method**
 1. How do we infer gene networks using single cell gene expression data?
 2. How do we evaluate if we've done a good job?

Questions for you:

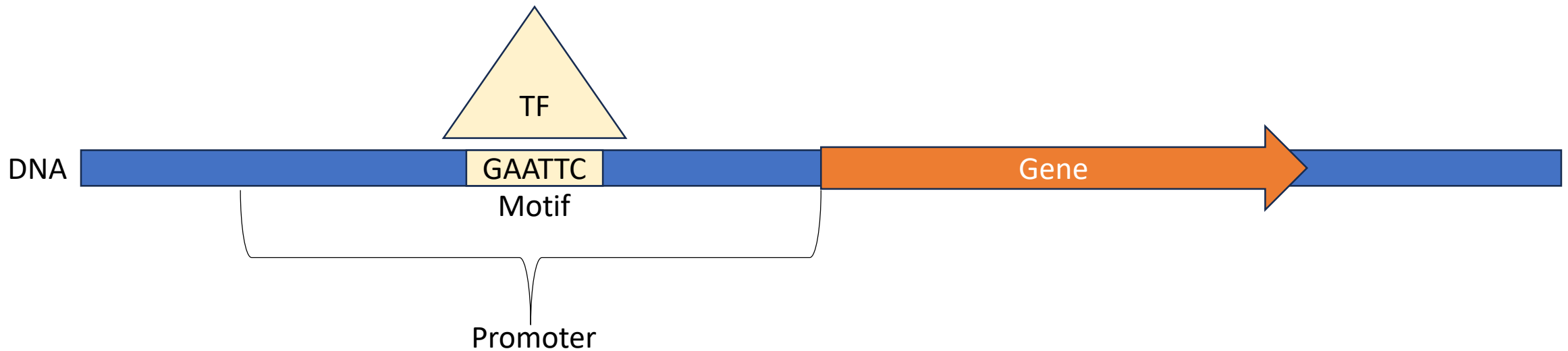
- What is your background?
- What do you hope to gain from this degree?

What is a transcription factor (TF)?



- Transcription factors bind to specific **motifs**. In plants (but NOT animals) these tend to be near the genes that they regulate.
- Transcription factor binding can lead to **increased or decreased** transcription

A fundamental question in biology:



Promoter sequence



Gene regulation

How do changes in DNA sequence in the promoter lead to changes in gene regulation?

What are transcription factor binding motifs like in plants?

(A) Position weight matrix

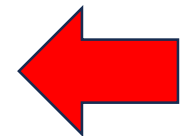


A	211	213	283	272	4	10	23	0	1	2	8	7
C	98	48	40	46	320	300	0	2	6	7	69	11
G	12	54	0	4	1	3	0	324	53	53	62	109
T	5	11	3	4	1	13	303	0	266	264	187	199

- Many transcription factors arose through historic gene duplication events (paralogs).
- Similar transcription factors are part of the same “family”
- Transcription factors in the same family tend to bind to similar motifs.

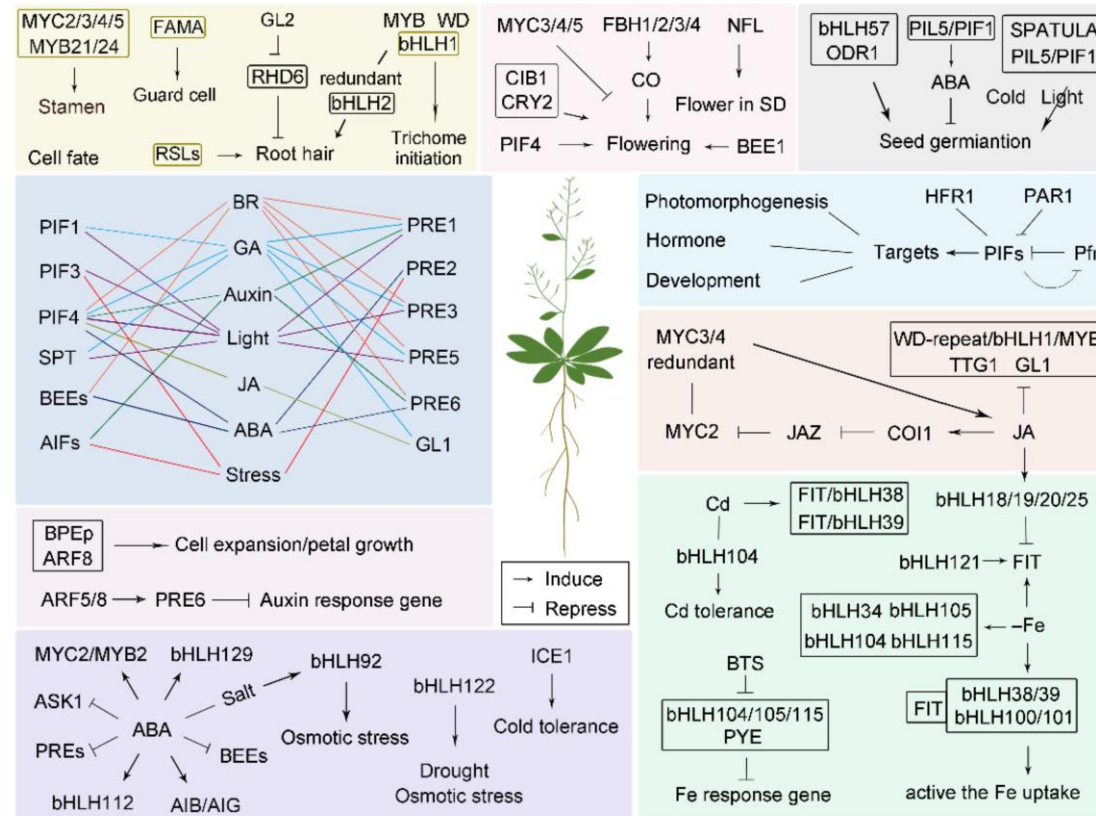
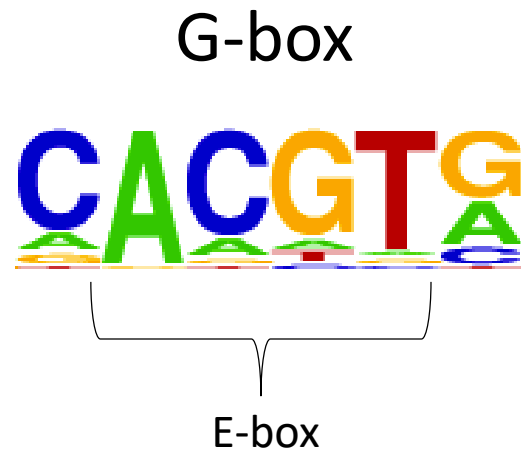
This is hard in plants because of the expansion of transcription factor families

Species	Myb_DNA-binding	z1-C2H2	z1-C3H4	CBFD_NFYB_HMF	AT_hook	PHD	Homeobox	HSF_DNA-bind	bZIP	ARID	SRF-TF	GATA	HLH	CBF	z1-B_box	E2F_TDP	Tub	CXC	AP2	WRKY	SBP	RWP-RK	CG-1	B3	z1-Dof	YABBY	EIN3	DUF573	GRAS	FLO_LFY	NAM	TCP
<i>Saccharomyces cerevisiae</i>	14	46	23	4	6	14	7	5	8	2	4	10	7	2																		
<i>Candida albicans</i>	16	52	27	7	2	15	3	7	7	2	2	9	11	3																		
<i>Schizosaccharomyces</i>	13	22	35	6	5	17	2	2	5	2	3	5	4	2																		
<i>Neurospora crassa</i>	16	75	35	5	16	14	6	2	9	3	2	6	14	2																		
<i>Fusarium graminearum</i>	12	95	30	5	20	15	10	3	10	2	2	6	16	2																		
<i>Magnaporthe grisea</i>	12	71	29	5	18	11	4	3	10	3	1	7	9	2																		
<i>Cryptococcus neoformans</i>	7	37	28	6	6	16	5	4	12	1	2	7	9	2																		
<i>Ustilago maydis</i>	11	33	28	7	12	13	6	4	5	2	2	8	12	2																		
<i>Encephalitozoon cuniculi</i>	9	11	15	5	1	3	10	2	1		1	2			1	2	1															
<i>Homo sapiens</i>	31	641	195	8	17	59	204	5	29	13	5	12	90	2	65	8	4	1				2										
<i>Gallus gallus</i>	29	279	133	6	20	47	126	4	21	10	2	7	70	1	14	8	3	2				1										
<i>Takifugu rubripes</i>	27	362	181	7	17	74	214	2	38	15	8	13	102	2	59	10	5	1				2										
<i>Ciona intestinalis</i>	18	166	69	6	7	36	78		11	6	2	2	39	1	9	5	1	1				1										
<i>Anopheles gambiae</i>	11	322	56	15	6	55	80	1	7	5	2	4	44	1	6	2	1	1				1										
<i>Drosophila melanogaster</i>	17	289	70	6	20	35	93	1	9	8	2	6	49	2	6	3	2	2														
<i>Caenorhabditis briggsae</i>	13	117	46	14	9	17	64	1	7	3	2	5	21	1	6	3	2	1														
<i>Caenorhabditis elegans</i>	15	168	90	6	13	24	83	1	9	4	2	12	42	2	15	4	2	1														
<i>Dictyostellum discoideum</i>	28	27	48	4	6	8	9		15	1	3	11		27	1			1														
<i>Entamoeba histolytica</i>	36	21	38	4		1	2	8	2		2	1		1	2																	
<i>Arabidopsis thaliana</i>	280	137	373	27	29	63	87	24	73	9	107	29	147	11	29	8	11	8	145	73	15	14	6	83	36	4	6	19	32	1	101	23
<i>Oryza sativa</i>	286	154	340	31	28	58	96	32	96	7	74	32	174	16	34	9	16	12	167	104	16	9	6	91	34	7	11	7	56	1	133	27
<i>Chlamydomonas</i>	22	10	37	4	13	32	3	2	8	2	1	6	8		1	2	3	1	15	1	11	6										
<i>Thalassiosira pseudonana</i>	35	2	53	6	3	22	5	24	5						3	3	7	9														
<i>Plasmodium falciparum</i>	7	8	22	2	1	5				1				3		1																
<i>Trypanosoma brucei</i>	1	28	18	1										2																		
<i>Giardia lamblia</i>	11	19	33	2		3				1				1	1					1												



What is a G-box?

- Many bHLH and bZIP transcription factors can bind to them and are involved in many processes



(a small selection of bHLHs, 10.3390/ijms22137152)

Mysterious G-boxes have been an issue for many Arabidopsis researchers

A **G-box**-like motif is necessary for transcriptional regulation by circadian pseudo-response regulators in **Arabidopsis**

TL Liu, L Newton, MJ Liu, SH Shiu, EM Farré - Plant physiology, 2016 - Am Soc Plant Biol

[HTML] W-box and **G-box** elements play important roles in early senescence of rice flag leaf

L Liu, W Xu, X Hu, H Liu, Y Lin - Scientific reports, 2016 - ncbi.nlm.nih.gov

[HTML] The regulation of the Z-and **G-box** containing promoters by light signaling components, SPA1 and MYC2, in **Arabidopsis**

SN Gangappa, JP Maurya, V Yadav, S Chattopadhyay - PloS one, 2013 - journals.plos.org

Role of the **G-box** element in regulation of chlorophyll biosynthesis in **Arabidopsis** roots

K Kobayashi, T Obayashi, T Masuda - Plant signaling & behavior, 2012 - Taylor & Francis

The **Arabidopsis** JAZ2 promoter contains a **G-Box** and thymidine-rich module that are necessary and sufficient for jasmonate-dependent activation by MYC ...

P Figueroa - Plant and Cell Physiology, 2012 - Jpn Soc Plant Physiol

Functional elements of the **Arabidopsis** Adh promoter include the **G-box**

WL McKendree Jr, RJ Ferl - Plant molecular biology, 1992 - Springer

Mutation of either **G box** or I box sequences profoundly affects expression from the **Arabidopsis** rbcS-1A promoter.

RG Donald, AR Cashmore - The EMBO Journal, 1990 - ncbi.nlm.nih.gov

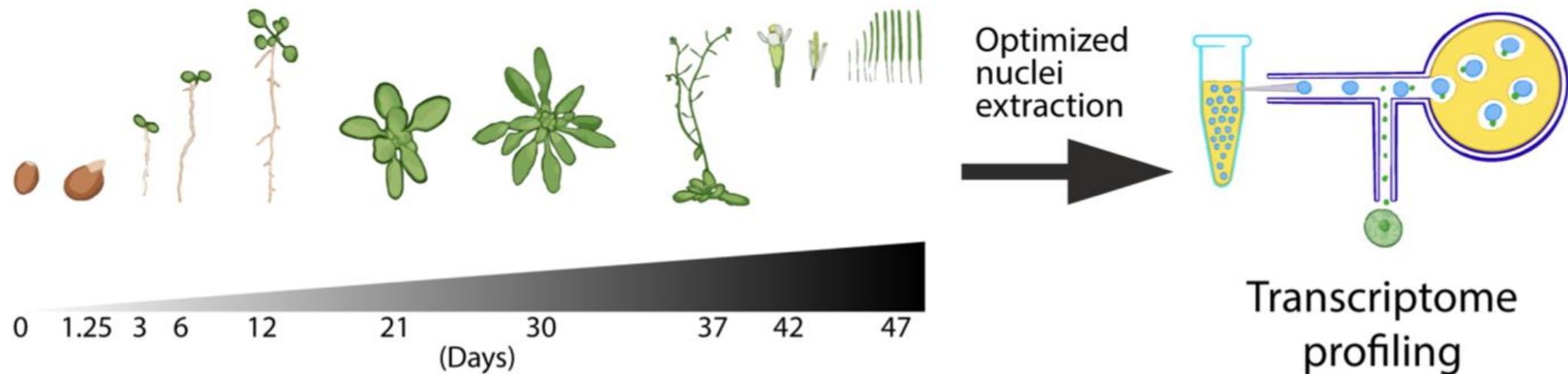
Ara-BOX-cis: the G-box regulatory network



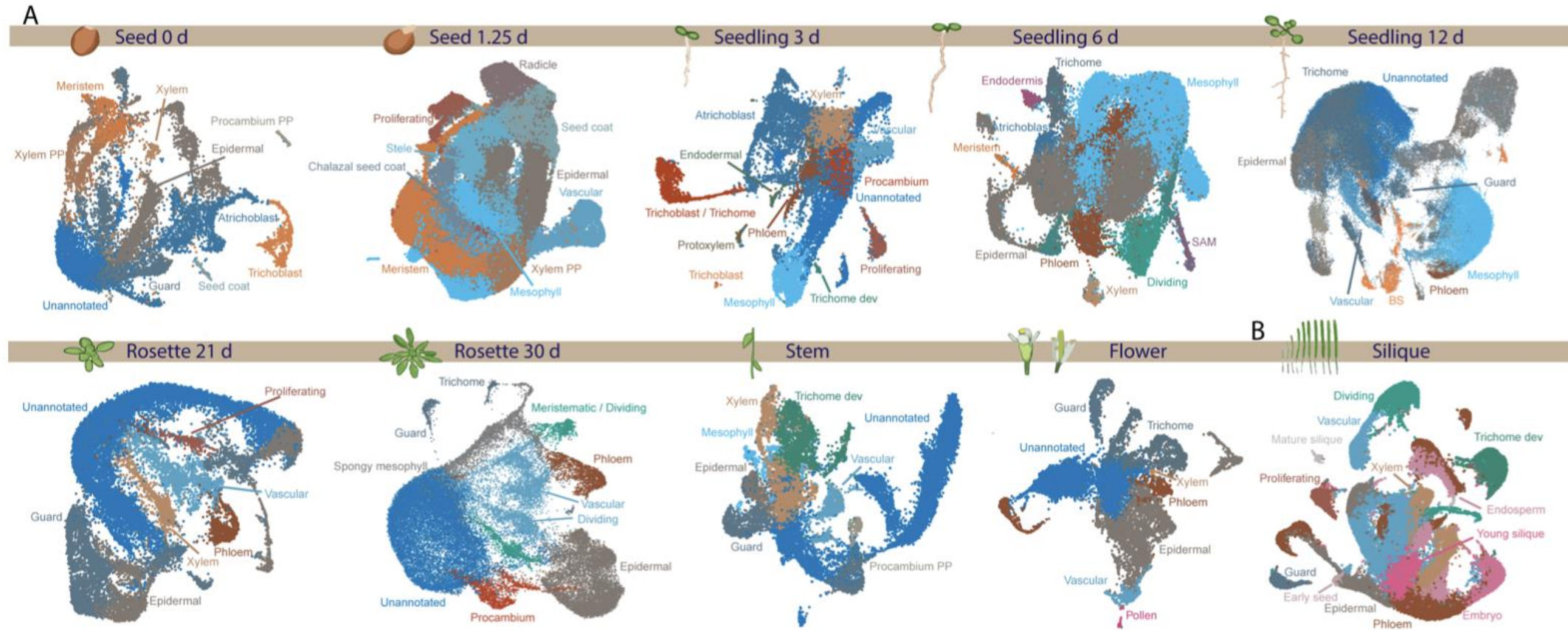
- We will discuss how this network was constructed in **Project Meeting 2**

Does Ara-BOX-cis need an update?

- Ara-BOX-cis was trained using a database of whole seedling RNA-seq
- Does it apply to different cell types?
- Does it apply to different developmental stages?
- New data available: a single nucleus developmental atlas (just released!)

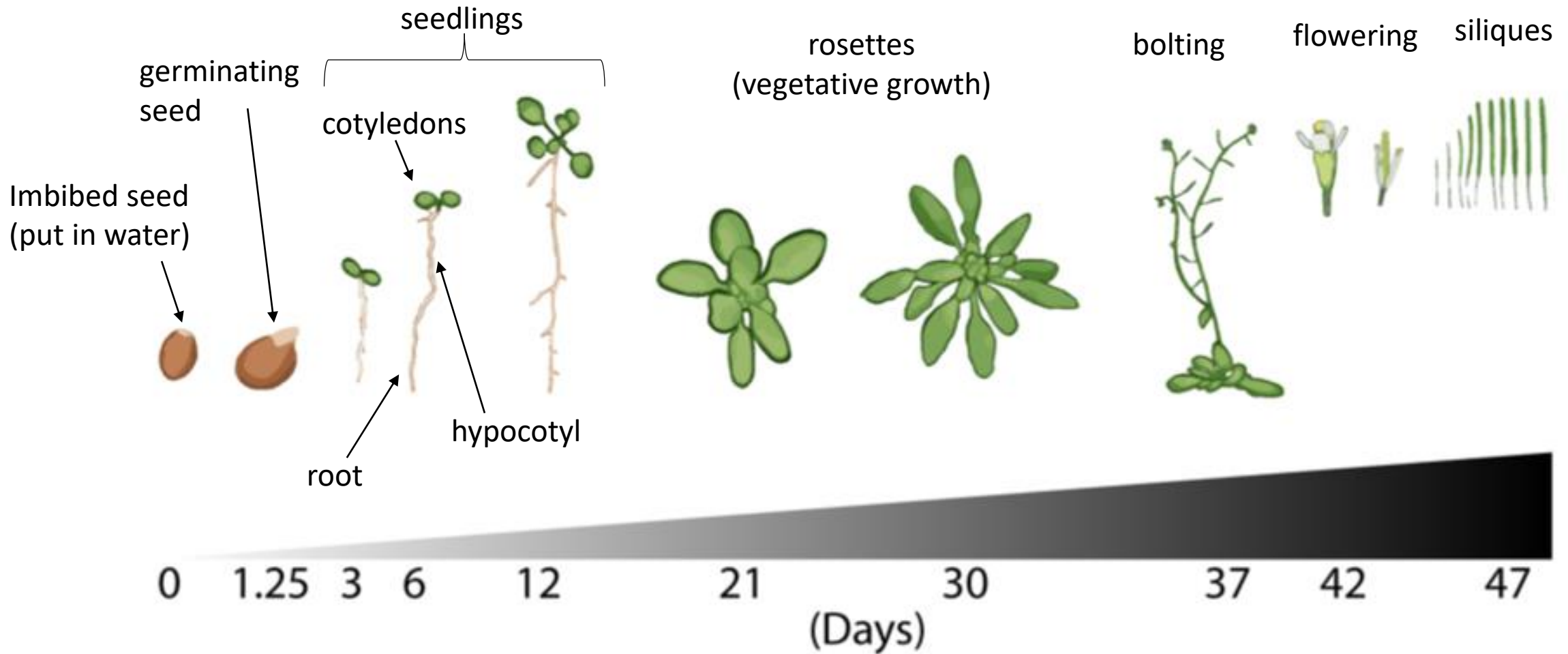


What data was produced by the single cell developmental atlas?

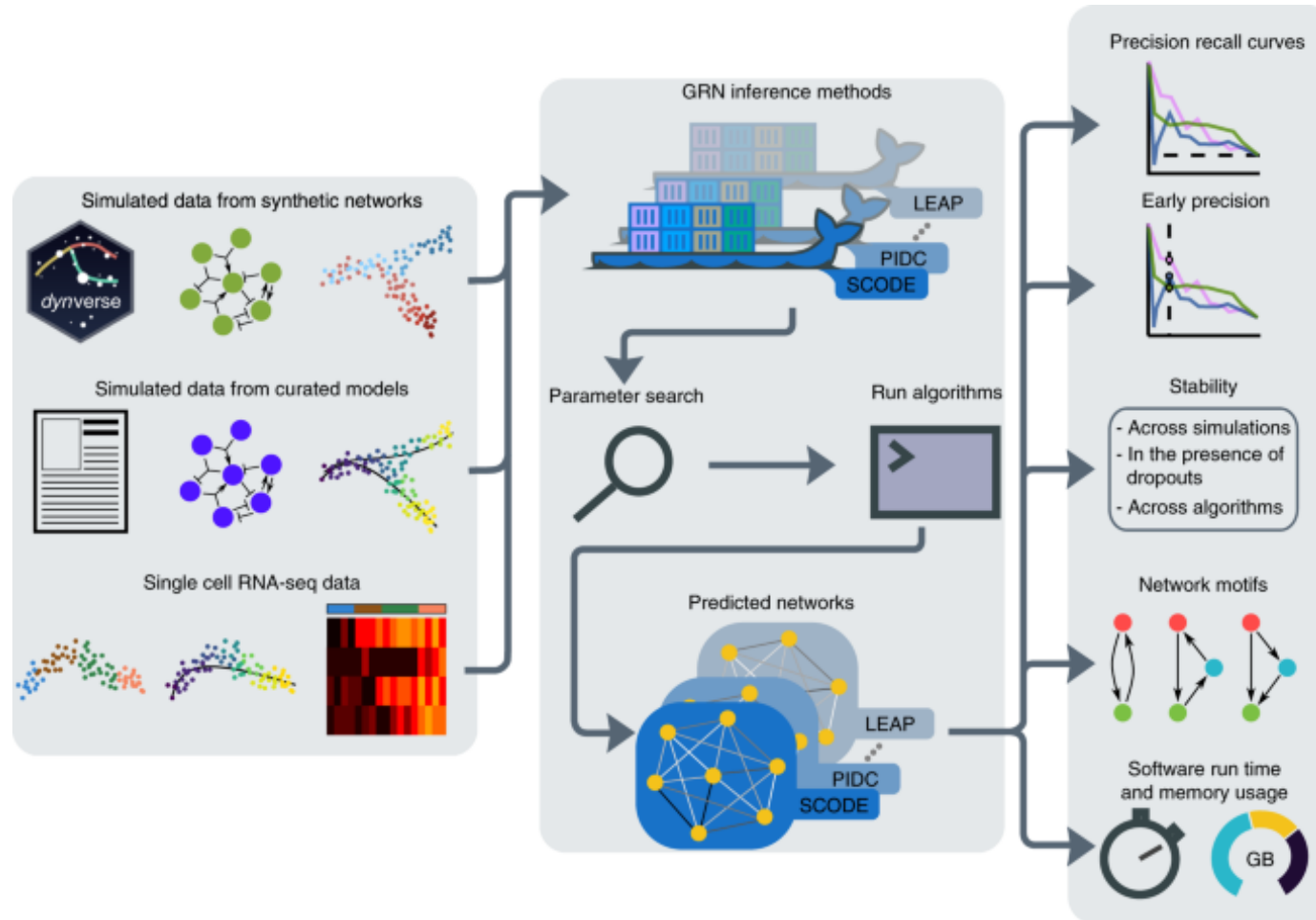


- We will discuss this dataset in depth in **Project Meeting 3**

Plant development overview:



How do we use single cell RNA-seq data to infer a gene regulatory network?



We will discuss how to build a network in **Project Meeting 4**

Aims of research project

- Is the existing Ara-BOX-cis G-box network consistent with the new single cell RNA-seq data?
- Do different G-box networks exist in different cell types or in different developmental stages?
- Can we construct Ara-BOX-cis version 2.0?

Outline of this term's material

1. Overview of project: what is the main aim?
2. Background on G-box regulatory networks
 1. What is a gene network?
 2. What is machine learning?
3. Background on the data set
 1. What is special about single cell RNA-seq?
 2. What are some exploratory analyses and visualisations we can do?
4. Background on method
 1. How do we infer gene networks using single cell gene expression data?
 2. How do we evaluate if we've done a good job?

Discussion questions:

- How does bioinformatics research differ from wet lab research?
- What are the similarities between bioinformatics and wet lab research?

Group work

- What makes a collaboration successful?
- What are some reasons why a collaboration might fail?