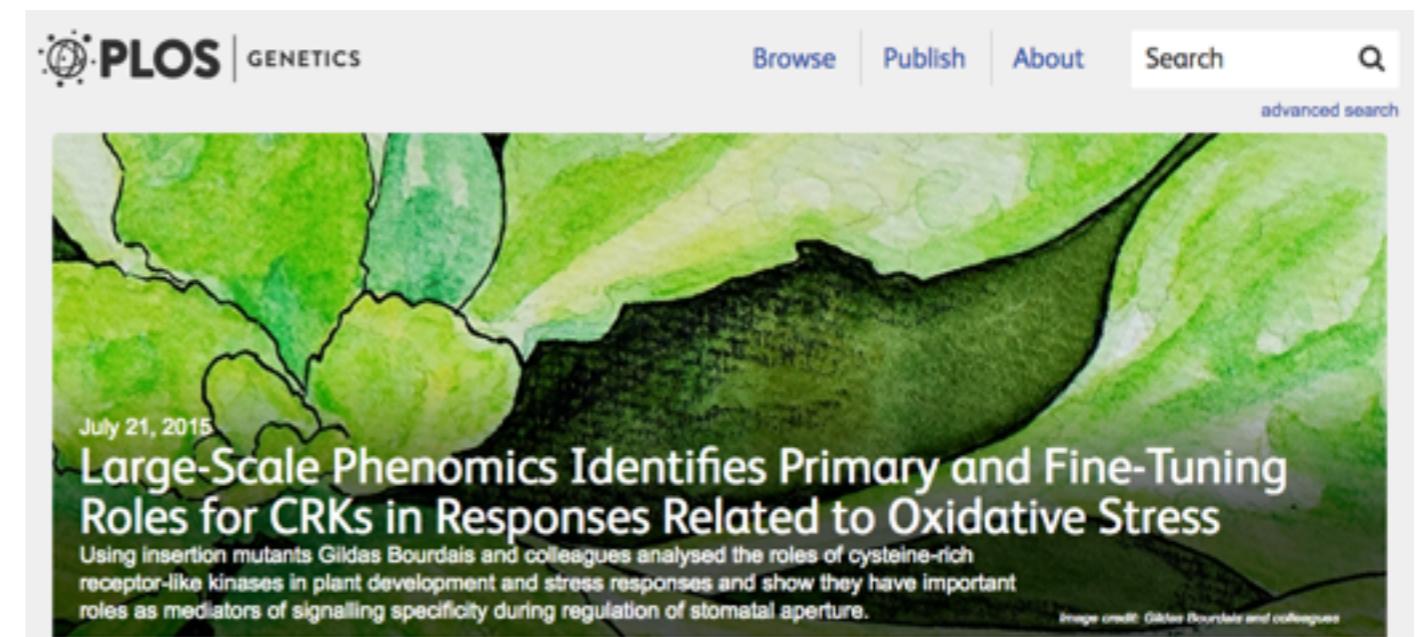


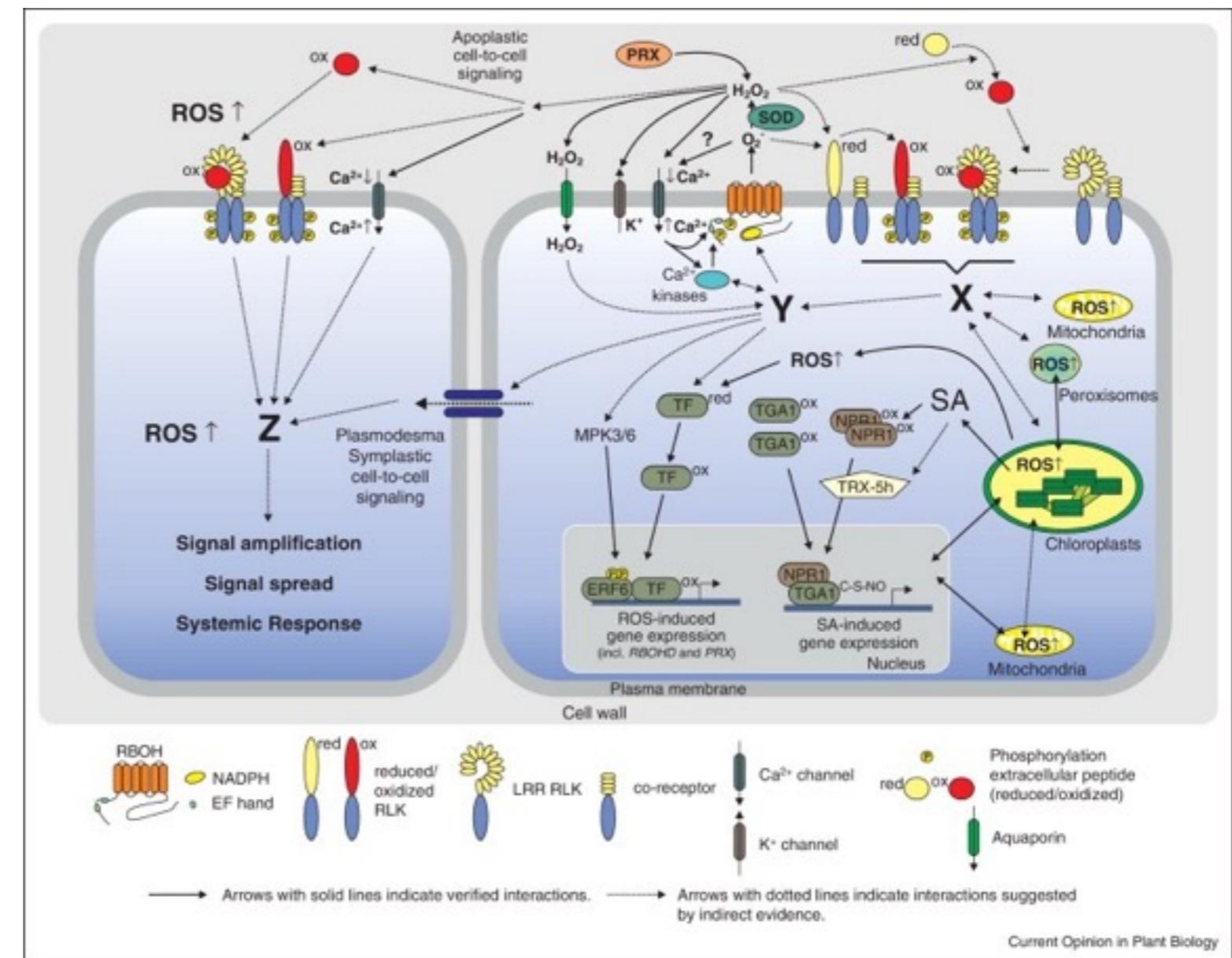
PhenomatoR: An R Package for the Analysis of Large-Scale Phenomics Data

Jarkko Salojärvi
Nanyang Technological University



ROS comes in many forms...and in many places

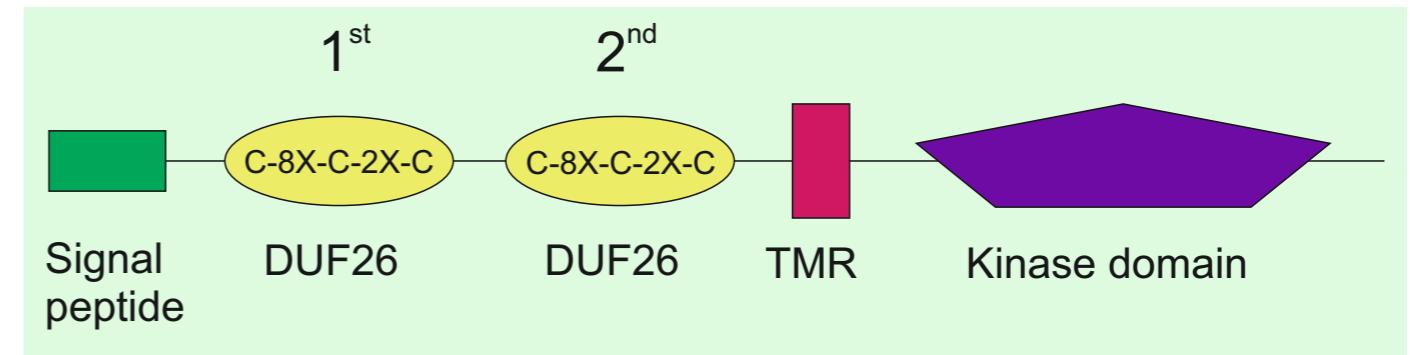
- ROS produced in various locations: mitochondrion, chloroplast, peroxisome, apoplast.
- Specific response to ROS depending on the source.
- Extracellular ROS perception still largely unknown.
- Many ROS signaling components involve oxidative modifications of cysteine residues.



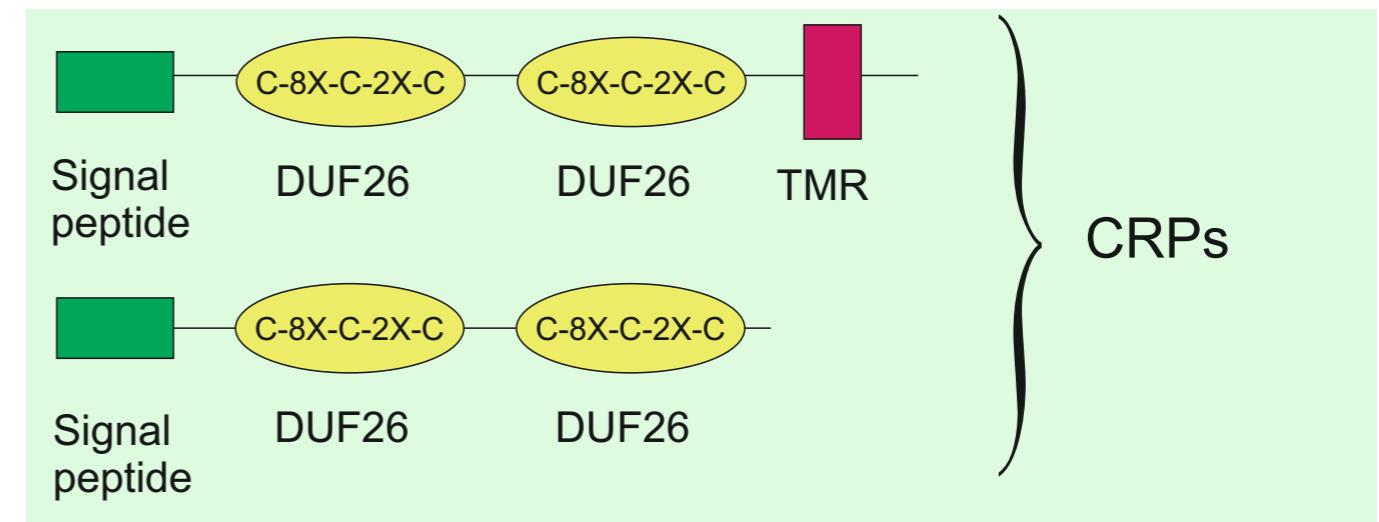
Wrzaczek, Brosché, Kangasjärvi: **ROS signaling loops – production, perception, regulation.**
Current Opinion in Plant Biology 16(5): 575–582, 2013.

Cysteine Rich Kinases (CRK) in *Arabidopsis thaliana*

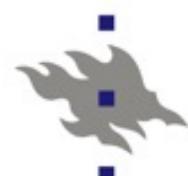
- Gene family of 44 members.
Consist of a signal peptide, 2 DUF26 domains, transmembrane domain and a kinase domain

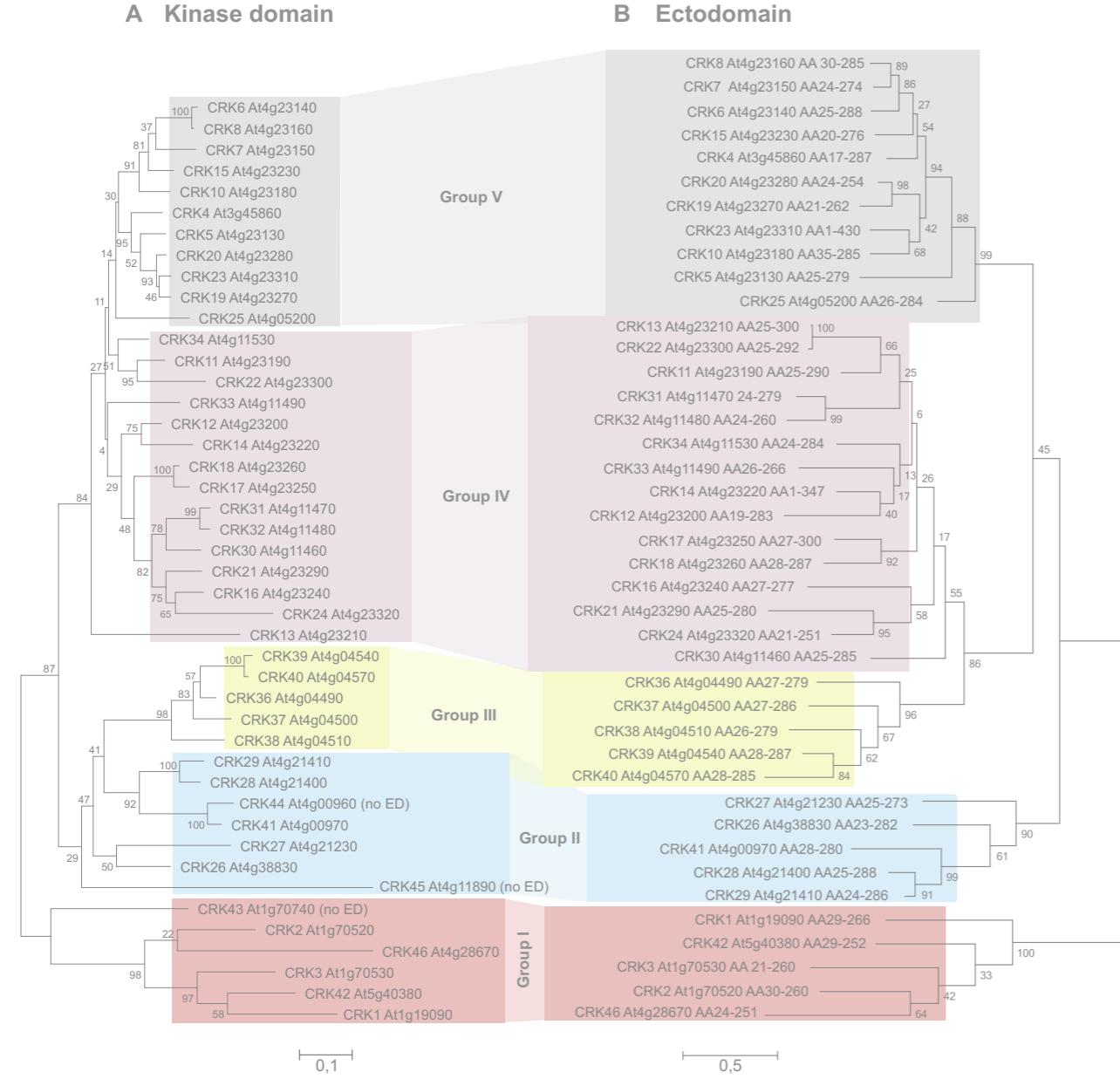


- DUF26 domain present also in:
 - 8 PDLPs (plasmodesmata localised proteins)
 - 50 CRSP (Cystein-rich receptor-like secreted proteins)

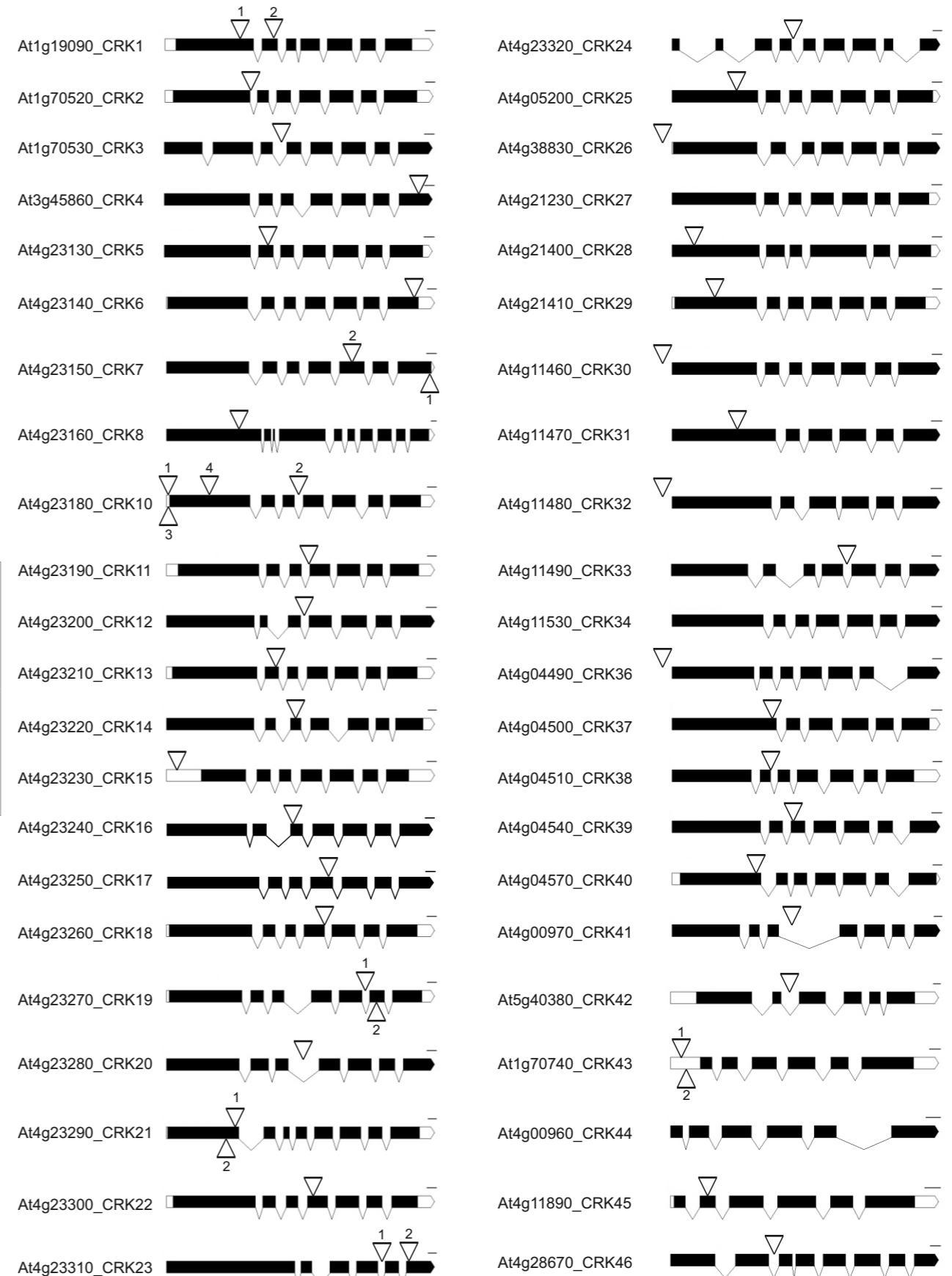


A. Vaattovaara





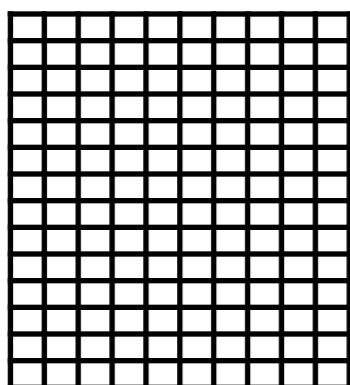
T-DNA insertion mutant collection for the crks



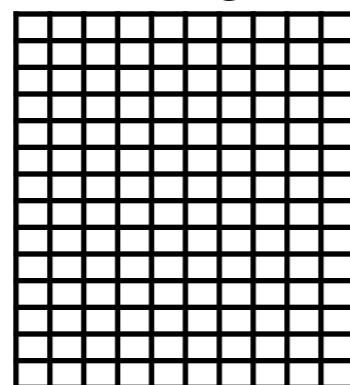
Large scale *phenomics* of CRK gene family

- Large-scale phenotyping of *crk* knockout mutants: Over 30 experiments characterizing developmental phenotypes, abiotic stresses, photosynthesis responses, pathogen susceptibility, and stomatal parameters.
- Highly heterogeneous, differing numbers of experiments + replicates

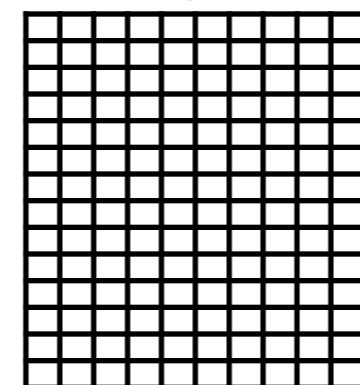
Abiotic stress



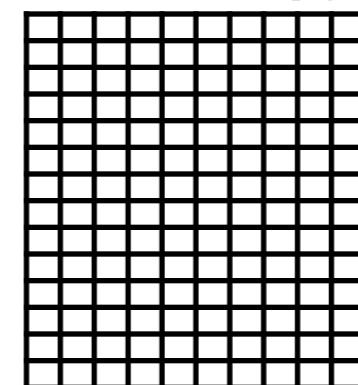
Pathogen



Photosynthesis



Microscopy

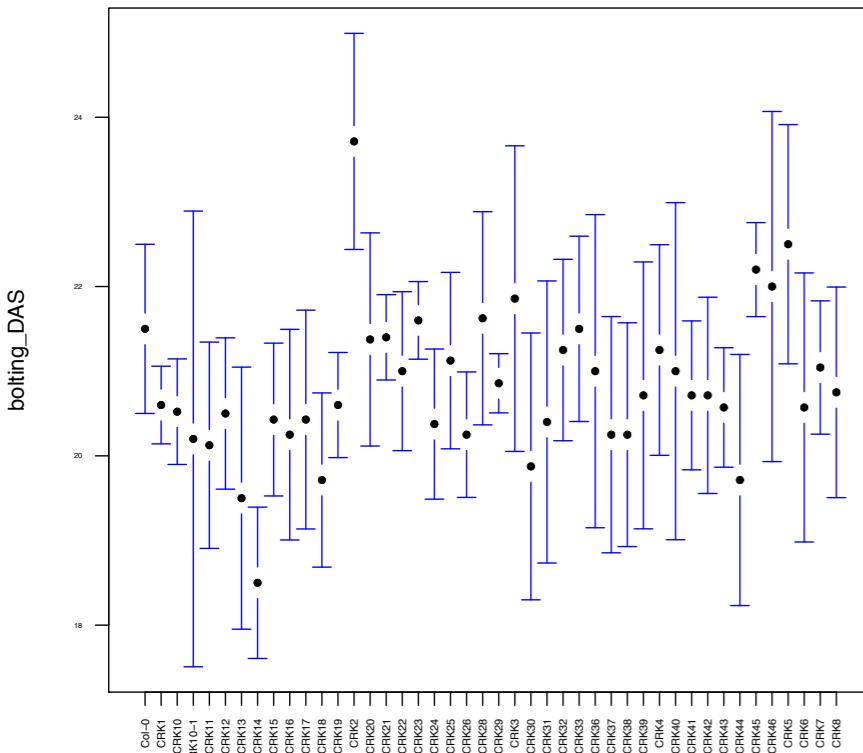


...

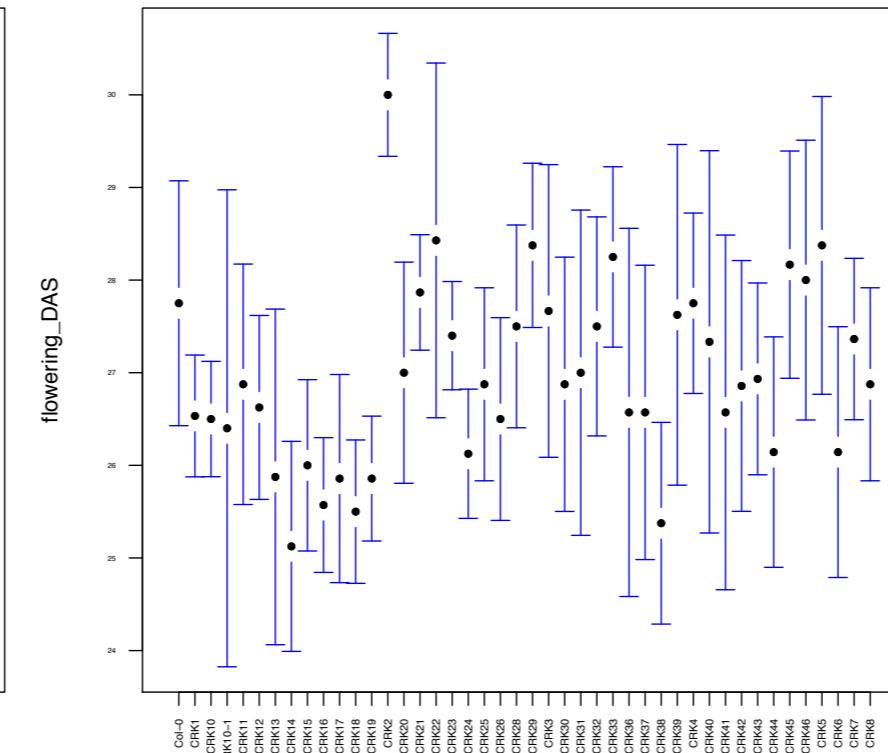
From low-throughput data...

- DAS = days after seeding

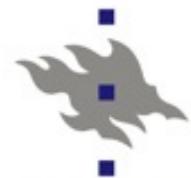
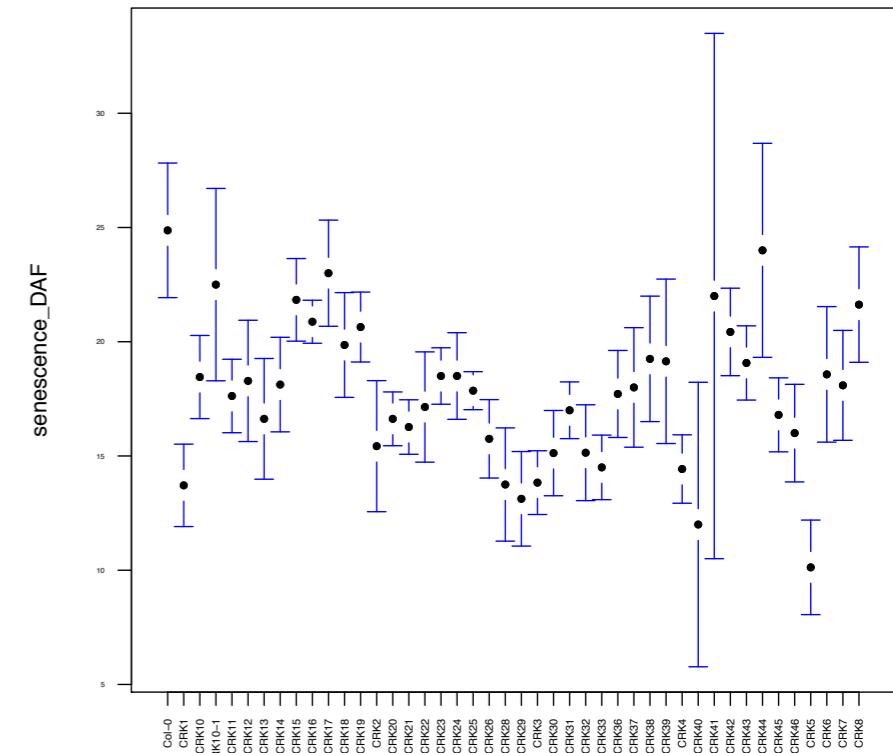
Bolting



Flowering

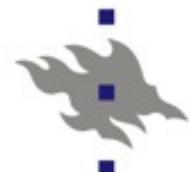
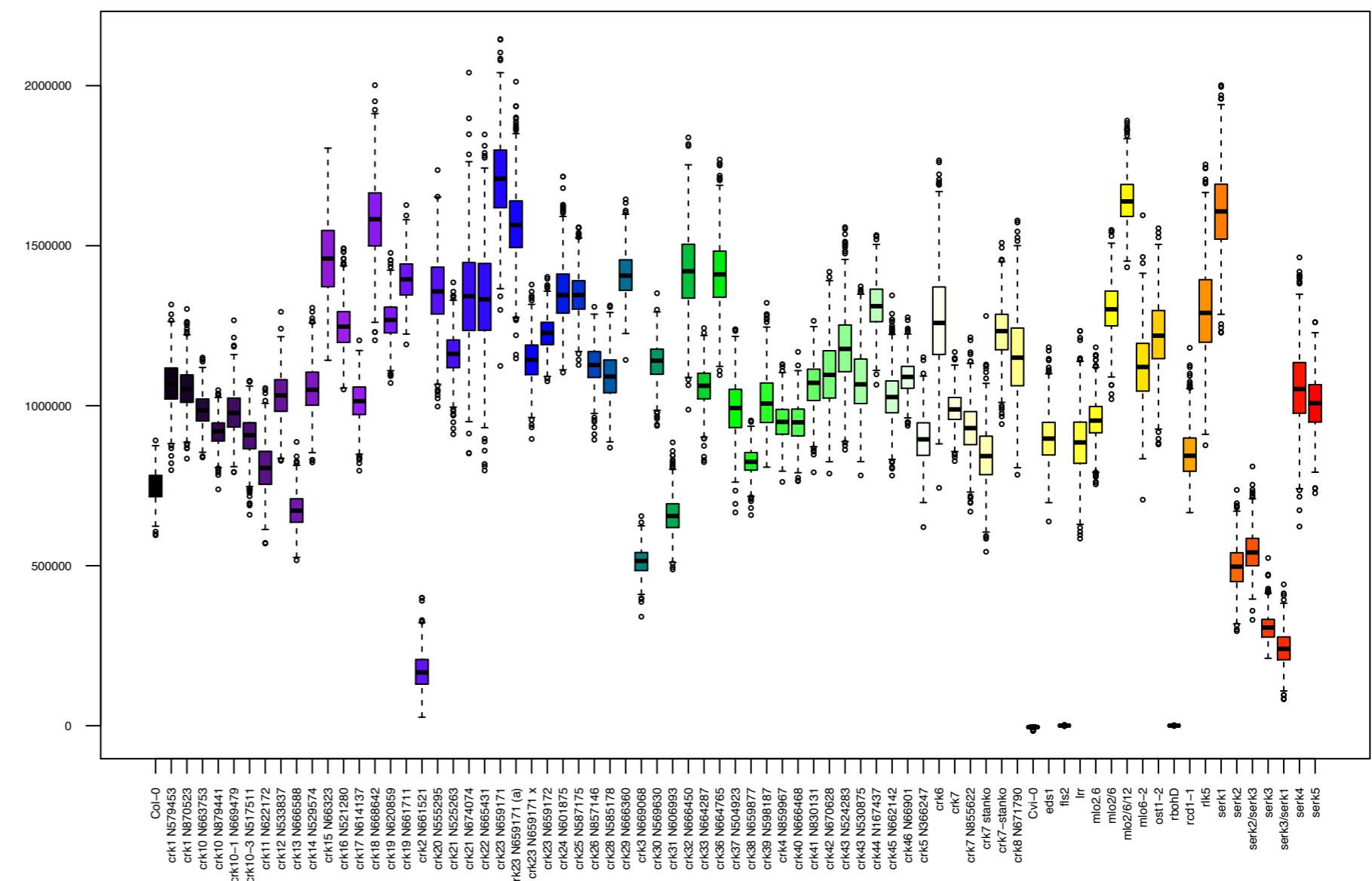
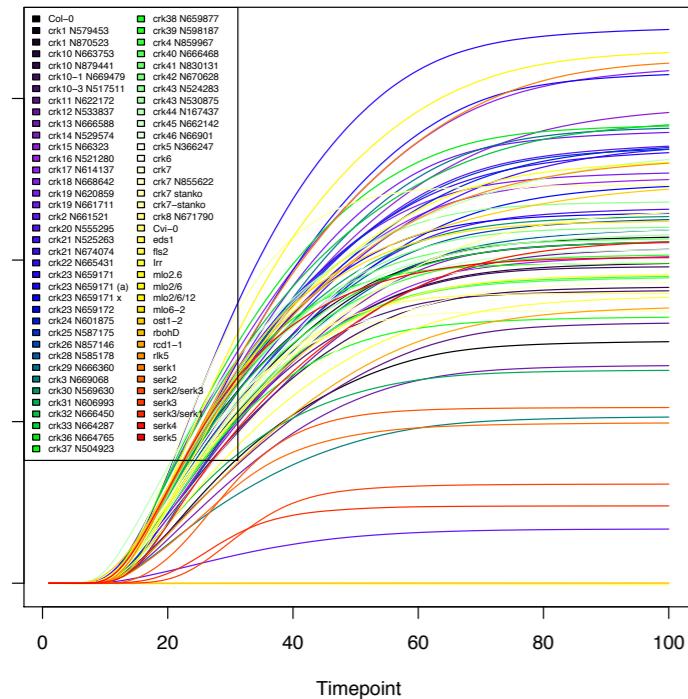


Senescence

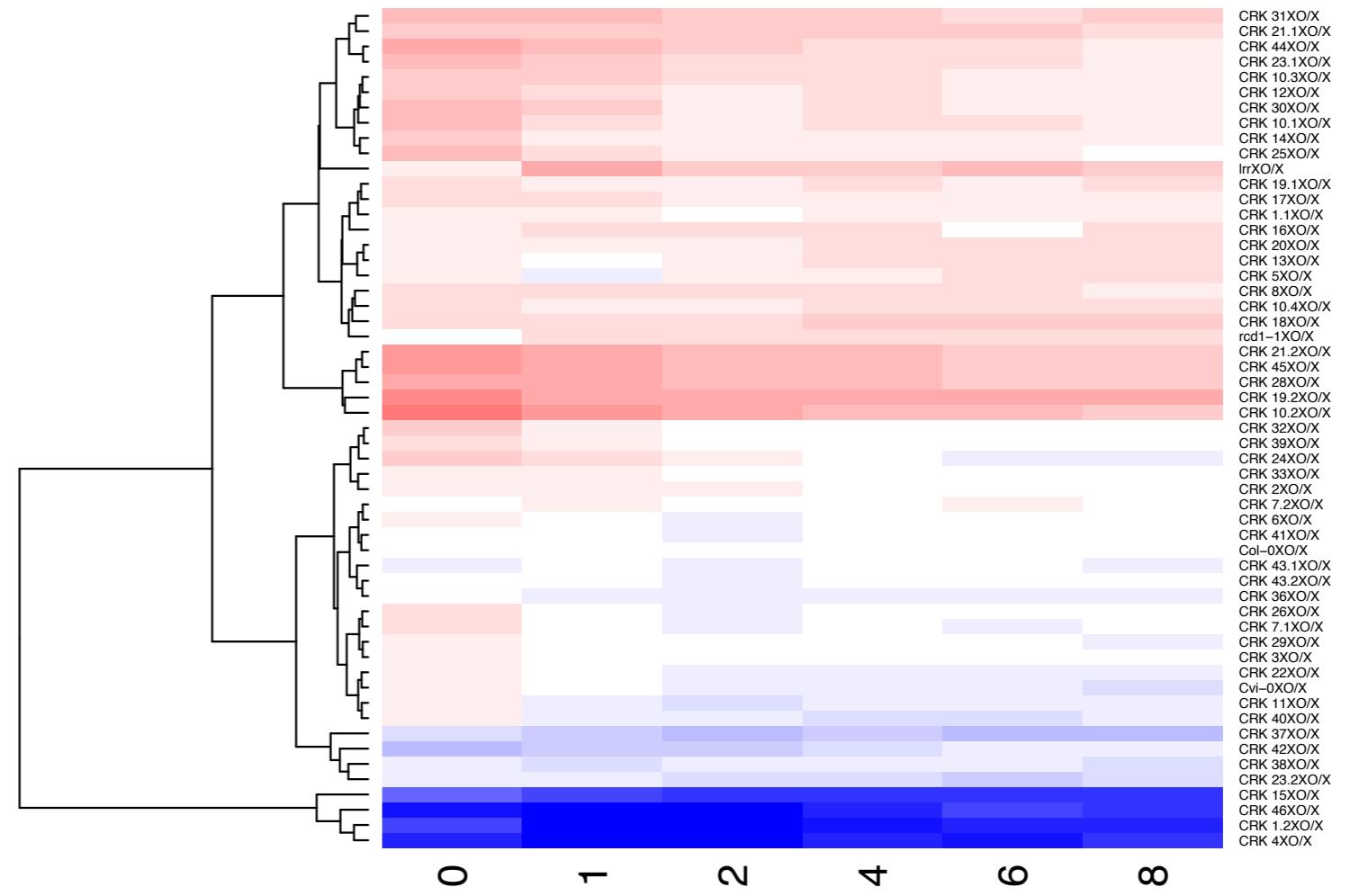
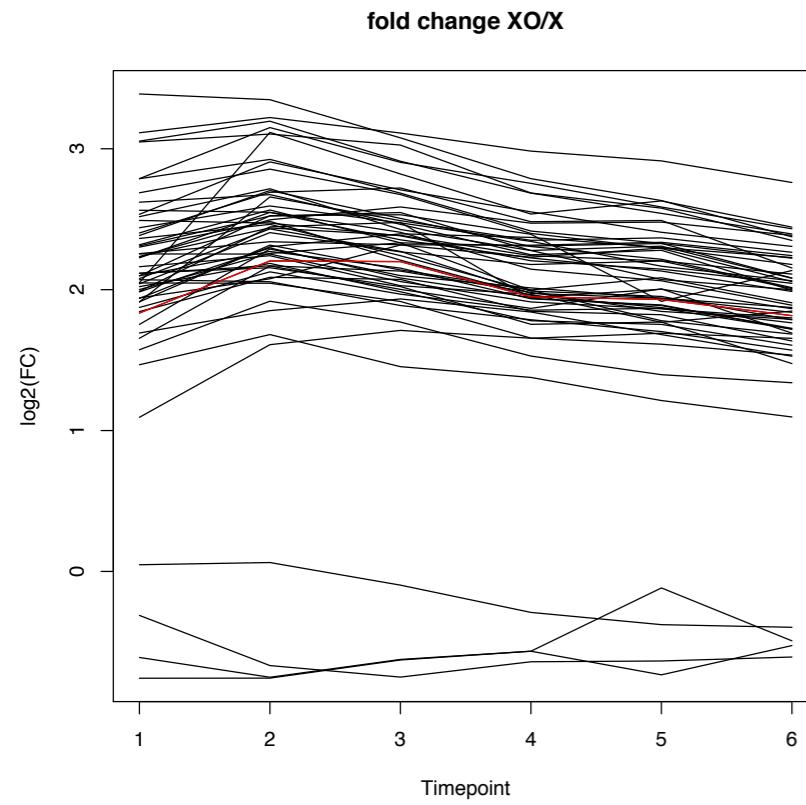
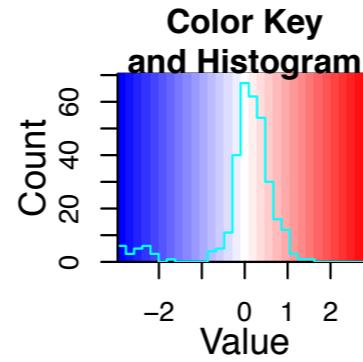


..to high-throughput microscopy

- Sample (1 leaf disk) at 100 time points (every 100s) for relative light unit.

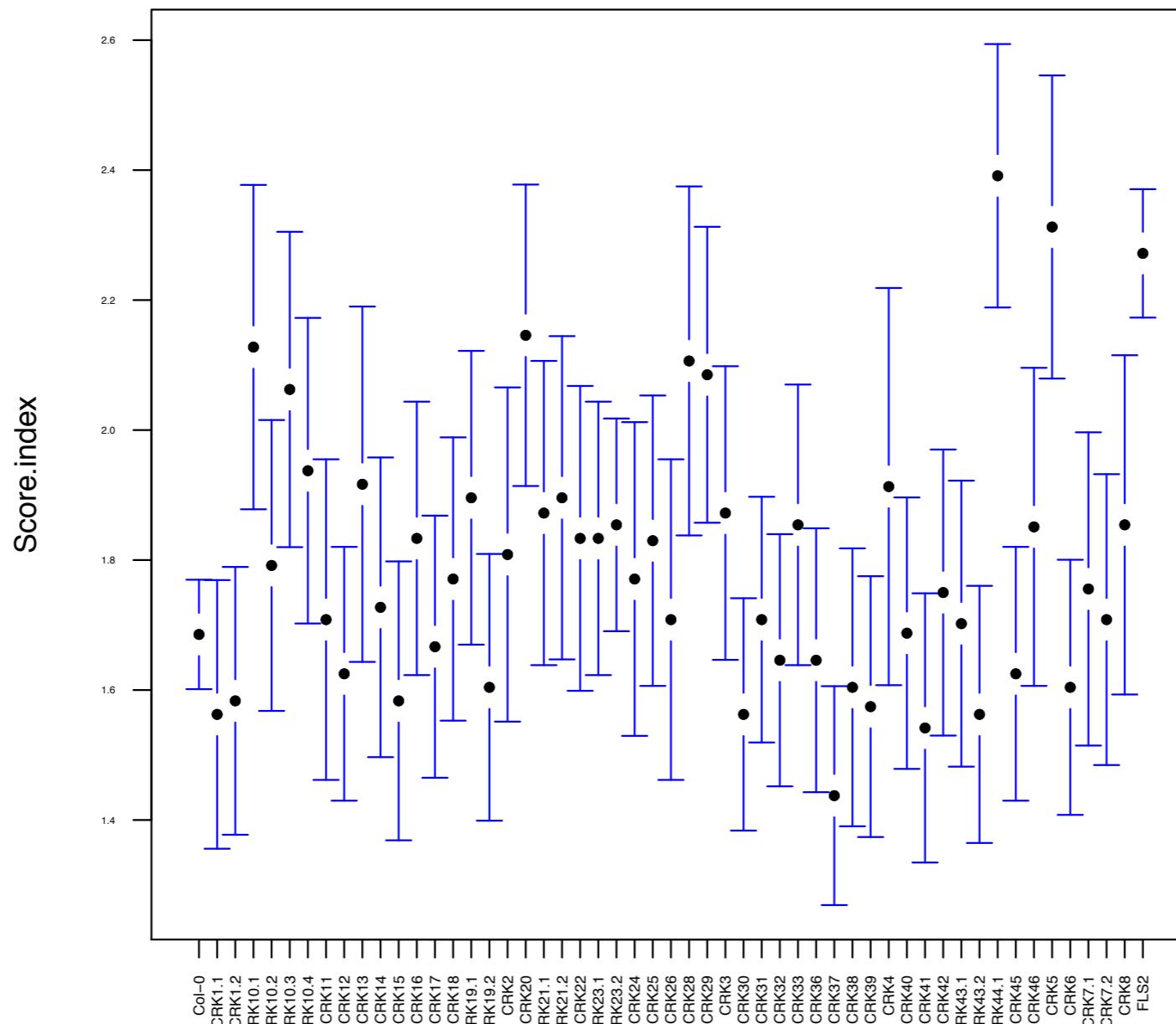


..and to time course data



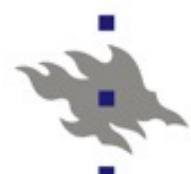
...and categorical data

- Infection with Pst-DC3000, scoring 1 (no symptom) - 4 (dead seedling)

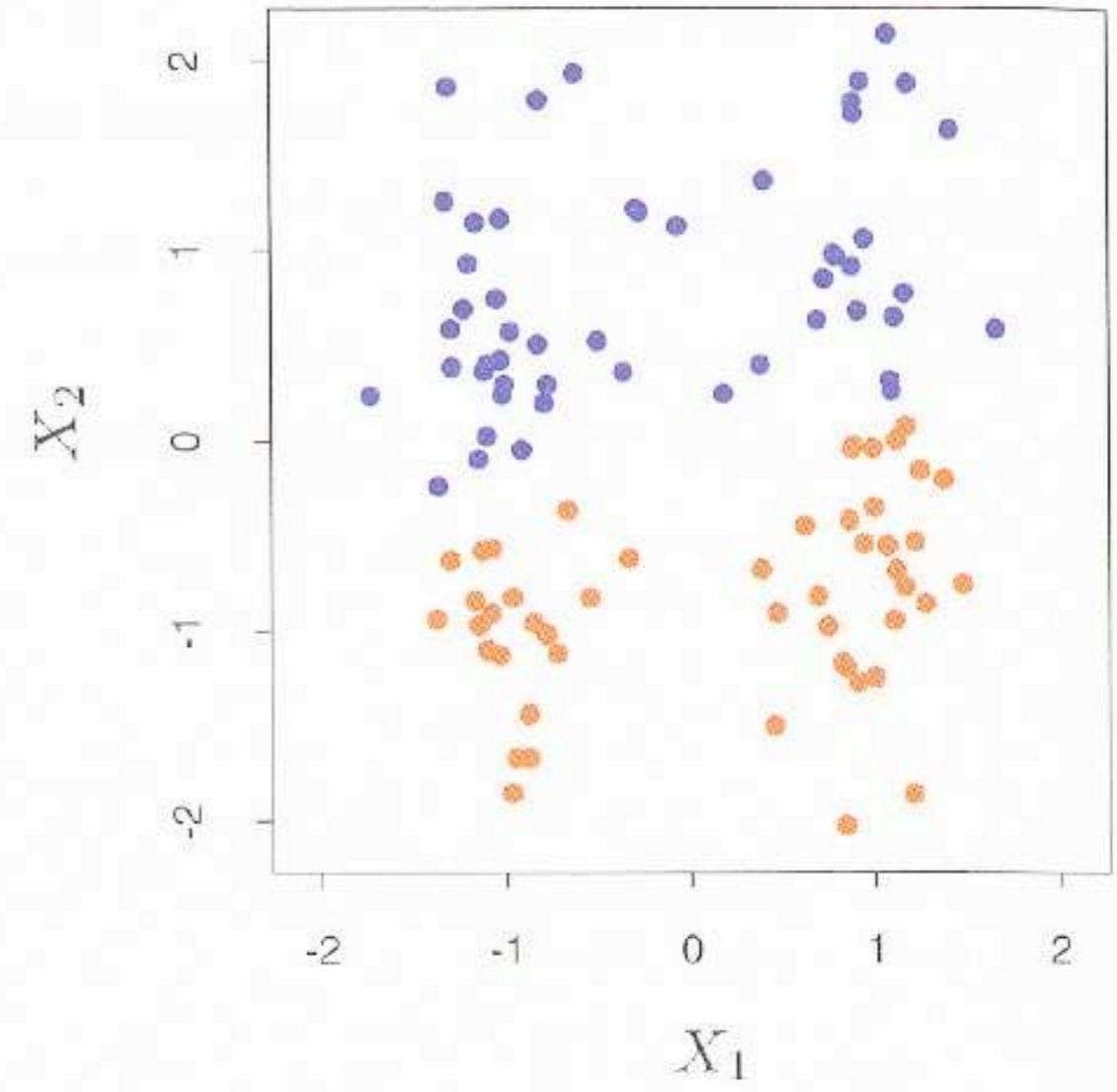
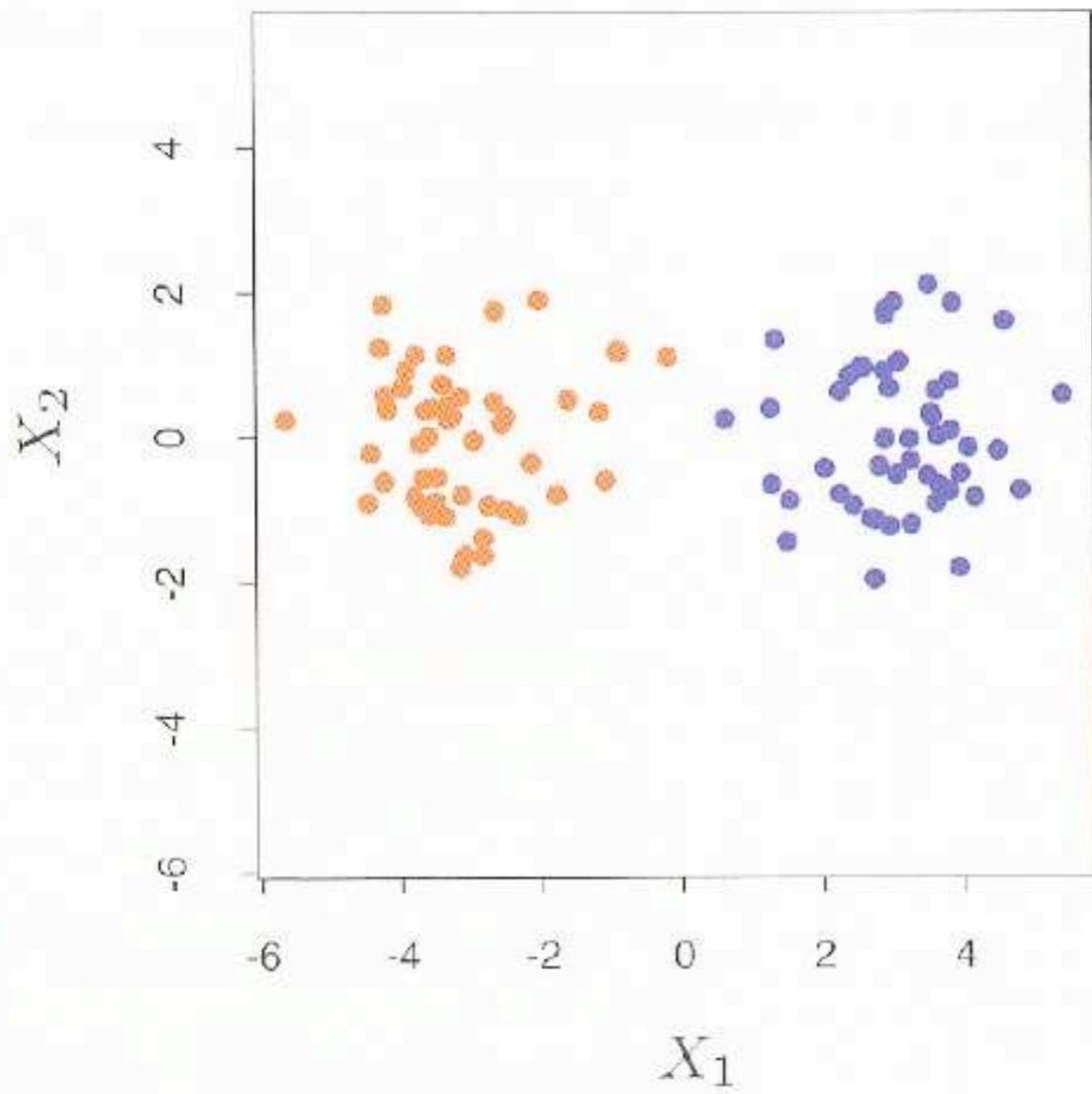


How to make data comparable: report outcomes of statistical tests in Z-scores

- Report all outcomes in terms of statistical distance from reference Col-0.
- BUT:
 - Different statistical tests have different null distributions.
 - The values of the test statistics are not comparable.
- Solution: approximate all test statistics with a Z-score (Gaussian distribution).
- Can be interpreted as p-value: $|Z| > 1.96$ corresponds to $p\text{-value} < 0.05$.
- Compared to p-value, tells also direction.



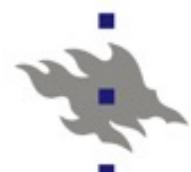
Why should I care?



In second plot, X_2 is scaled to have same variation as X_1 .

Further improvements

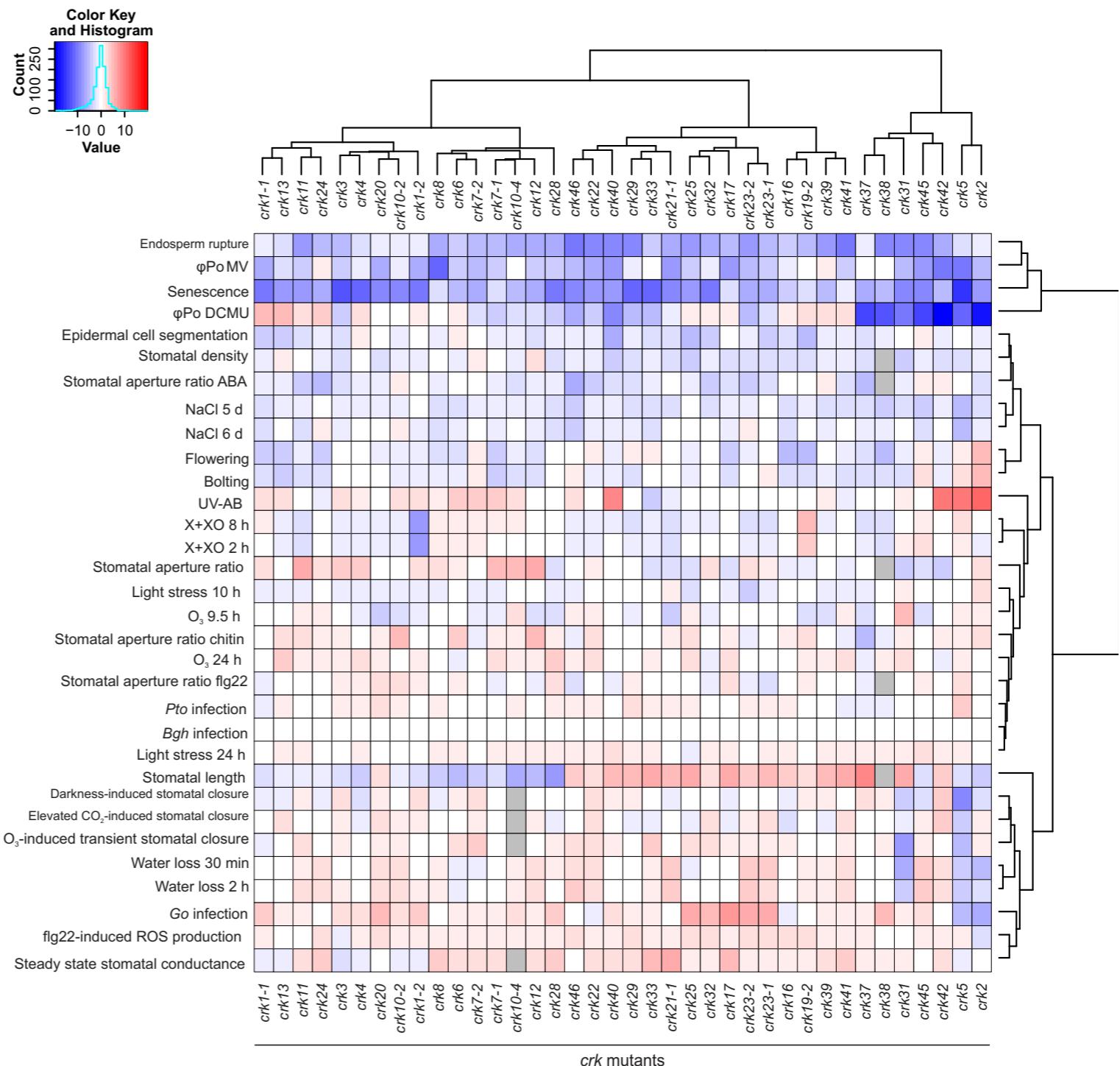
- With enough repetitions, even tiny non-relevant differences get statistical significance.
 - High-throughput methods can automatically measure 100's of values (e.g. stomata size in leaf discs).
 - Low-throughput measurements such as disease index are quantified only once per plant, and the value is indicative.
- Leads into biased estimates, since high-throughput measurements will have more statistically significant differences.
- Solution: downsample the experiments to the same number of replicates.
 - Carry out statistical testing with this set to get Z-score.
 - Bootstrap to provide error estimate to the Z-score.



Large-scale phenomics: summary

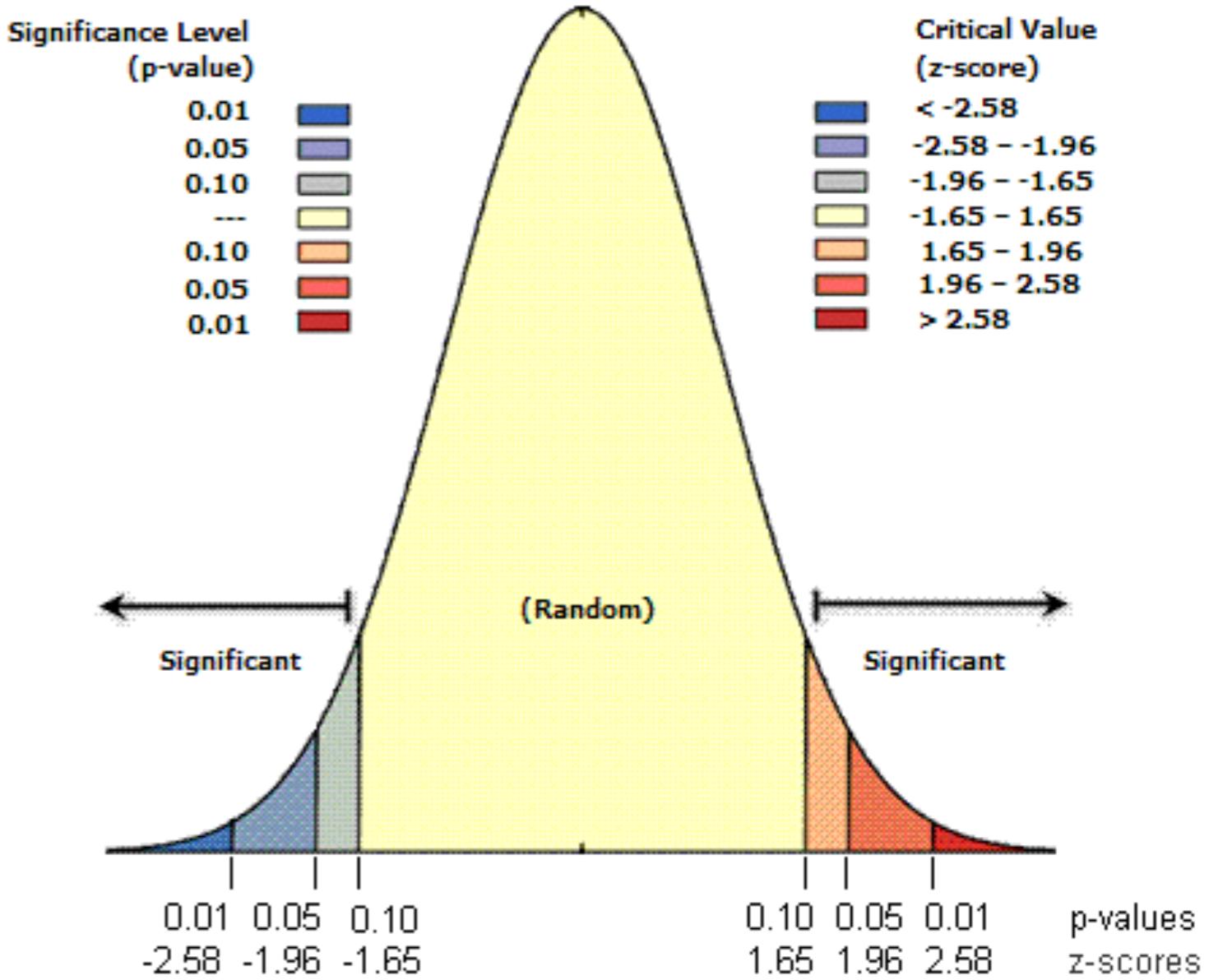
- Each experiment summarized to a Z-score that can be compared across experiments.
- Bootstrap sampling to get same number of repeats (N=15).
- Can be interpreted as p-value:

$|Z| > 1.96$ corresponds to p-value < 0.05.



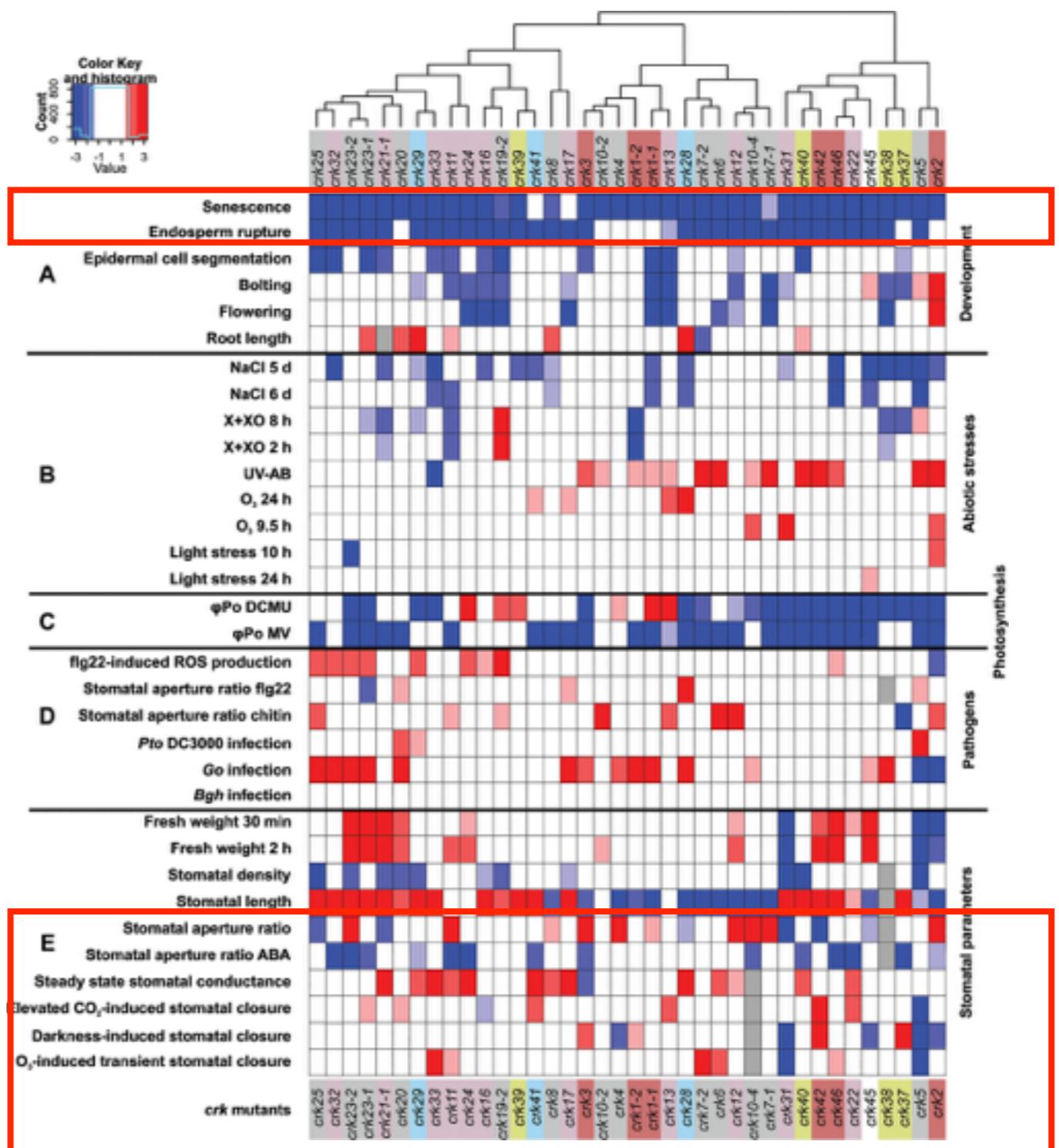
Multiple statistical testing..?

- Even though the heatmap does not show p-values, the Z-scores have a one-to-one mapping.
- FDR correction can be done by converting to p-values, carrying out FDR correction and converting back to corresponding Z score.



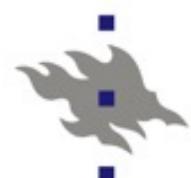
Significant differences wrt Col-0 wild type

- All crk knockout mutants show significant phenotypical differences wrt Col-0
- Whole family has similar developmental phenotype.
- From high sequence similarity within the family, one would have expected great level of redundancy.
 - Now: each CRK has a specific role.
 - Several antagonistic roles.



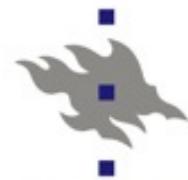
Can I do similar analysis..?

- R package under development: PhenomatoR:
 - **<https://github.com/jsalojar/PhenomatoR>**
- Functions for:
 - Estimating or converting different data and statistical test types into z-scores.
 - Downsampling/bootstrapping the data.
 - Carrying out FDR correction for the estimated values.
 - Can be visualized as standard heatmap in R.
- Will become available by the end of February.



Works with Photon System Instruments data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Measuring Date	Measuring Time	Experiment ID	Round Order	Tray ID	Tray Info	Plant ID	Position	Plant Name	Plant Info	PID	Camera Position	AREA_PX	AREA_MM	PERIMETER_PX	P
2	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI001	A1	Col-0	Col-0	RGB2		41	3412	95,589711	529,1858582
3	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI002	A2	B	32-1	RGB2		41	5605	157,02823	665,7544007
4	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI003	A3	C	66-17	RGB2		41	5612	157,22434	515,4284989
5	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI004	A4	A	1C-2	RGB2		41	5119	143,41258	605,8132752
6	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI005	A5	D	86-2	RGB2		41	5920	165,85319	636,2985566
7	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI006	B1	Col-0	Col-0	RGB2		41	4757	133,27088	601,2691193
8	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI007	B2	A	1C-2	RGB2		41	3225	90,350767	514,8010819
9	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI008	B3	D	86-2	RGB2		41	4781	133,94326	596,3990617
10	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI009	B4	D	86-2	RGB2		41	4475	125,37044	509,9726548
11	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI010	B5	Col-0	Col-0	RGB2		41	5067	141,95576	638,1980515
12	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI011	C1	C	66-17	RGB2		41	4710	131,95414	445,7888886
13	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI012	C2	B	32-1	RGB2		41	6082	170,39174	636,1980515
14	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI013	C3	A	1C-2	RGB2		41	2842	79,620738	438,8599564
15	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI014	C4	Col-0	Col-0	RGB2		41	4421	123,85759	565,3279938
16	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI015	C5	C	66-17	RGB2		41	4389	122,96109	432,0731599
17	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI016	D1	B	32-1	RGB2		41	4994	139,91061	641,9970414
18	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI017	D2	A	1C-2	RGB2		41	3417	95,729789	416,7178208
19	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI018	D3	D	86-2	RGB2		41	5814	162,88352	661,5950212
20	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI019	D4	B	32-1	RGB2		41	4168	116,76961	565,6711396
21	14/06/2016	10:30:43		22	11	FUPS_VIA_0	FUPSVI020	D5	C	66-17	RGB2		41	4803	134,55961	505,3868684
22	14/06/2016	10:31:31		22	11	FUPS_VIA_0	FUPSVI021	A1	C	66-17	RGB2		40	3499	97,704358	389,646753
23	14/06/2016	10:31:31		22	11	FUPS_VIA_0	FUPSVI022	A2	Col-0	Col-0	RGB2		40	3817	106,58403	474,3158005
24	14/06/2016	10:31:31		22	11	FUPS_VIA_0	FUPSVI023	A3	A	1C-2	RGB2		40	3863	107,86851	515,529004
25	14/06/2016	10:31:31		22	11	FUPS_VIA_0	FUPSVI024	A4	B	32-1	RGB2		40	3260	91,030639	500,8010819
26	14/06/2016	10:31:31		22	11	FUPS_VIA_0	FUPSVI025	A5	Col-0	Col-0	RGB2		40	4163	116,24557	511,7716447
27	14/06/2016	10:31:31		22	11	FUPS_VIA_0	FUPSVI026	B1	D	86-2	RGB2		40	4695	131,10087	616,6417023



...and Lemnatec data

	A	B	C	D	E	F	G	H	I
1	Row No	Results	Analysis Label	Partial Analysis	Analysis Time Stamp	2nd Moment Principle Axis Large Abs	2nd Moment Principle Axis Small Abs	2nd Moment Principle Axis Large Norm	2nd Moment Principle Axis Small No
2	1	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	129107800	82978850	0.1647136	0.1058631
3	2	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	69206810	31723160	0.1805746	0.08277213
4	3	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	52579460	43545780	0.1541951	0.1277028
5	4	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	76174970	42361230	0.1749245	0.09727627
6	5	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	190979200	153951400	0.13065	0.1053191
7	6	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	103109700	82891540	0.138592	0.1114163
8	7	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	66834700	54000290	0.1391361	0.1124175
9	8	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	0	0	0	0
10	9	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	170951000	117135600	0.1601083	0.1097062
11	10	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	166054900	117572000	0.1423537	0.1007908
12	11	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	98873550	76508970	0.1611102	0.124668
13	12	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	167584800	148358900	0.1327829	0.1175496
14	13	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	74975230	58705890	0.1457963	0.1141591
15	14	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	149015300	125475600	0.1207534	0.1016782
16	15	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:26	82127770	68272420	0.1299553	0.1080312
17	16	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	5664966	1336061	0.2878818	0.06789584
18	17	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	2681928	666714.8	0.2860465	0.07110984
19	18	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	1106080	382333.5	0.2522512	0.08719448
20	19	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	1042913	380550.5	0.2563517	0.09354068
21	20	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	7398280	3018918	0.2247034	0.09169175
22	21	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	5369519	1431676	0.2754698	0.07344855
23	22	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	3405504	685676.4	0.3127185	0.06296386
24	23	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	0	0	0	0
25	24	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	5865896	849332.8	0.3419119	0.04950599
26	25	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	7301732	1478320	0.3297023	0.06675203
27	26	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	4191000	753606.1	0.3233797	0.05814862
28	27	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	6315529	1961621	0.2554744	0.07935107
29	28	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	1802395	531606.6	0.2364379	0.06973609
30	29	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	11697930	2151464	0.3028495	0.05569956
31	30	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:23	5387091	1326895	0.3290808	0.08105595
32	31	ObjectListROI	Exp1_soil_15112017	FALSE	11/15/17 9:19	132877.2	16531.66	0.289062	0.0359631

More example data sets needed...!

Bioinformatics for Molecular Biology & Genomics



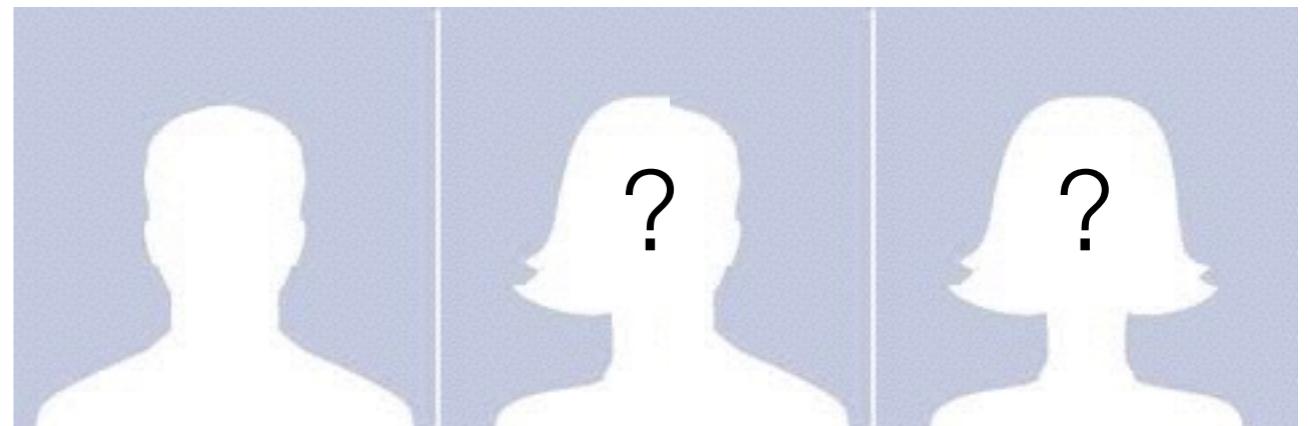
Thanks!



Maja Ilievska, Cui Wang (postdoc)



Postdoc needed..



Nicholas Cho

jarkko@ntu.edu.sg



Centre of Excellence of Academy of Finland
Molecular Biology of Primary Producers

