

Transcriptome analysis and *de novo* motif discovery predict putative nitrogen responsive regulatory DNA motifs in potato (*Solanum tuberosum* L.)

José Héctor Gálvez López

M.Sc. Plant Science Candidate

MonBUG May 4 2016

Martina Strömvik, Ph.D. (Supervisor)

Helen Tai, Ph.D. (Co-supervisor)



McGill



Solanum tuberosum L. (Potato)

In 2012, Canada's potato production was worth \$1.12 billion CAD.

(Source: Statistics Canada)

Importance of Nitrogen (N) Fertilization:

N Insufficiency: unsuitable for food industry; lower yield.

N Excess: lower potato quality; environmental damage.

No added N (0 kg N ha^{-1})



With added N (180 kg N ha^{-1})



Photos by: Helen Tai

Solanum tuberosum L. (Potato)

Is there a genetic response to N availability?
What mechanisms regulate potato's N response?

Genomic characteristics:

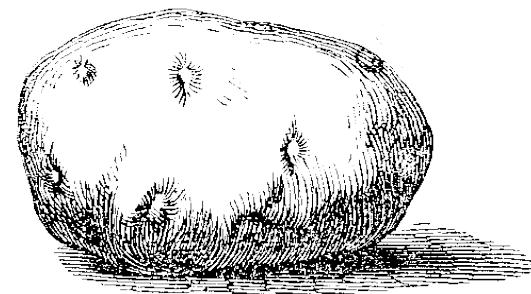
Chromosome number = **12 = n**

Ploidy = **4n** (tetraploid)

Genome size ≈ **723Mb**

Number of protein-coding genes ≈ **39,301**

(The Potato Genome Sequencing Consortium, 2011; Sharma *et al.*, 2013)



Transcriptomic characteristics:

High-confidence transcripts = **22,704**

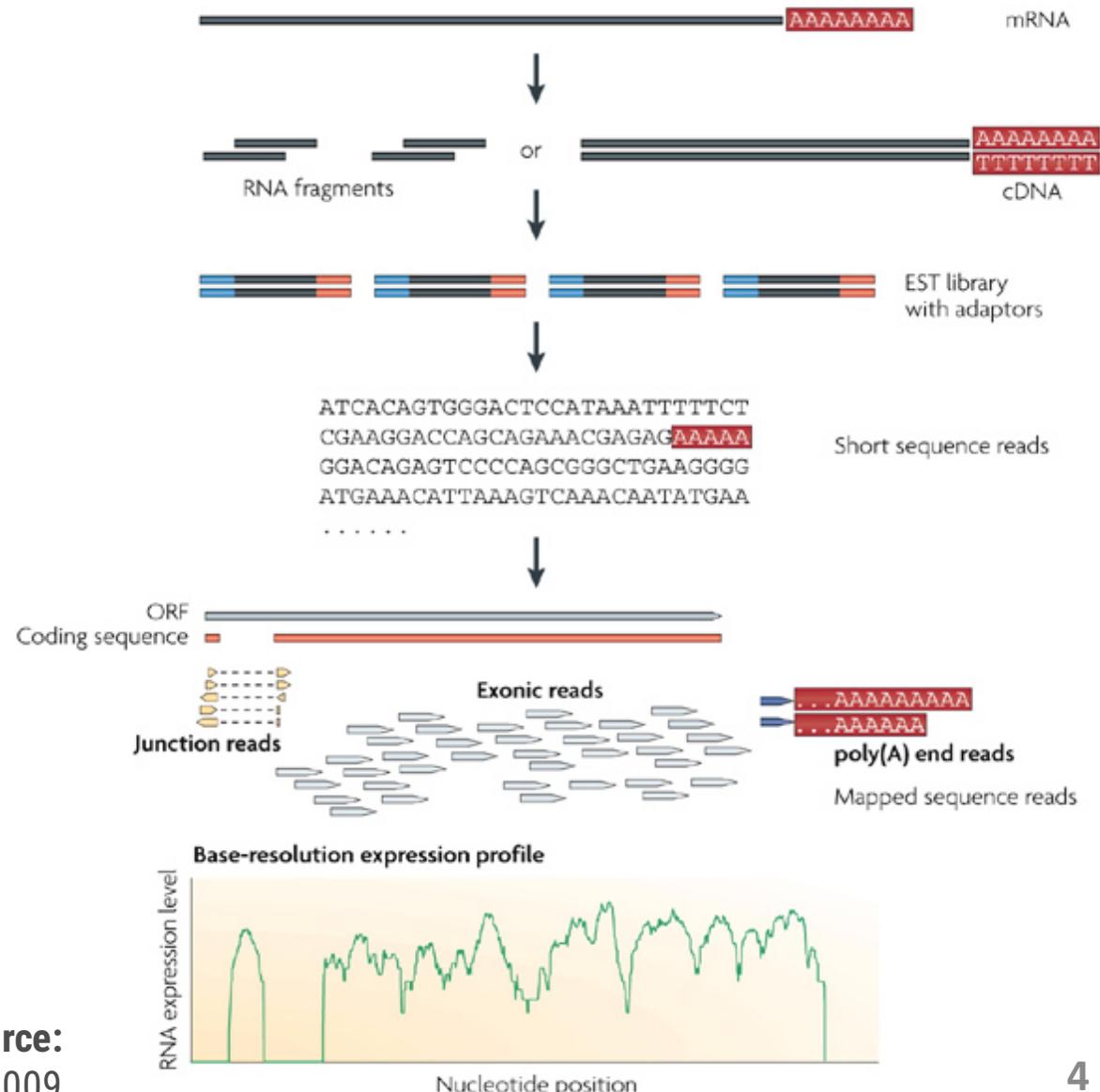
- 17% Unknown function
- 8% Only found under stress

(Massa *et al.*, 2011)

Differential Gene Expression in Potato

Available techniques:

- RT-qPCR
- Arrays
- nCounter
- **RNA-seq**



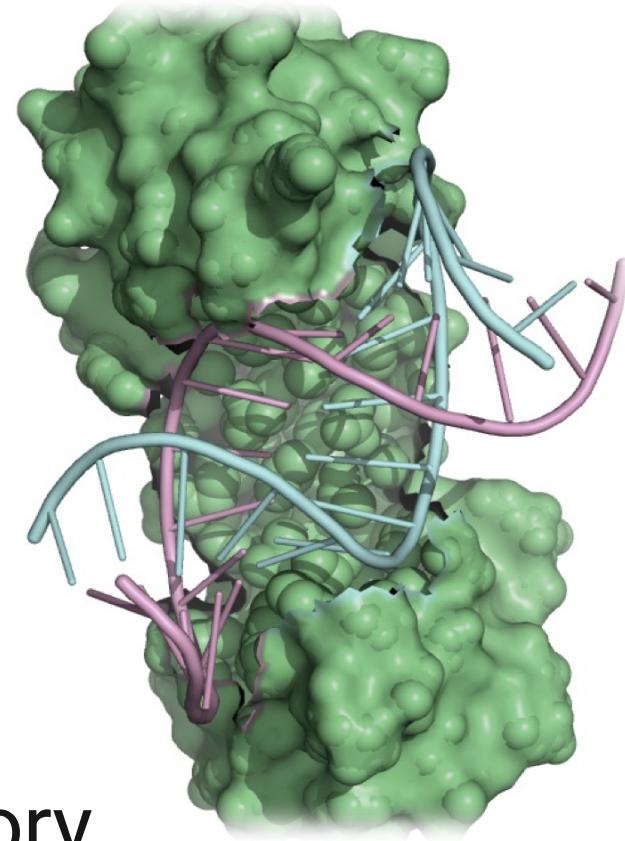
Source:
Wang et al., 2009

Regulatory motifs

Transcription factors recognize and bind to specific sequences or '**motifs**'.

“The promoters of co-expressed genes are likely to share **common regulatory motifs** and are potentially regulated by a common set of transcription factors.” (Bi, et al., 2007)

Finding these common regulatory motifs can help us understand the regulation of genes.



N Regulatory motifs

N regulatory motifs identified in other plants:

- ***Arabidopsis thaliana*** (Konishi *et al.*, 2010)
 - Nitrate Related *cis*-Element (NRE)
 - Discovered in the upstream region of the *Nitrate Reductase* (NIR) gene.
 - Necessary and sufficient for nitrate-activated gene transcription
 - Also found in the upstream region of NIR genes in spinach, tobacco, rice, maize and sorghum.
- **Maize (*Zea mays L.*)** (Liseron-Monfils *et al.*, 2015)
 - Eight overrepresented motifs discovered in the upstream regions of nitrogen-related genes.

Hypotheses

1. When exposed to long term differences in N supplementation, **potatoes of three cultivars have a common set of N responsive genes.**
 - Differentially expressed genes N responsive will mostly belong to pathways related with N metabolism.
2. Differentially expressed genes in *S. tuberosum* under N-stress **have overrepresented motifs** in their upstream promoter regions.

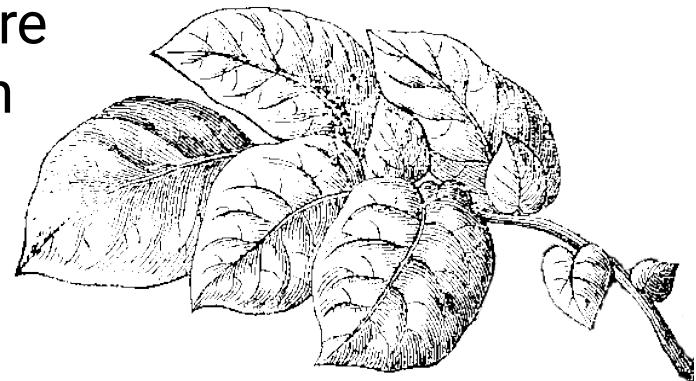


Objectives

1. Use RNA-seq data to **identify N responsive genes** in potatoes of three different cultivars.
 - Analyze the differentially expressed genes to find overrepresented pathways.
2. Analyze the upstream promoter regions of differentially expressed genes to **detect overrepresented motifs**.
 - Install Seeder and other *de novo* motif discovery programs for use within a **High Performance Computing** (HPC) environment.
 - Develop automated strategies to deal with redundancy in motif discovery results.

Experimental design

Planted in the experimental field of Agriculture and Agri-Food Canada, Fredericton Research and Development Centre (Helen Tai, et al. 2012)



48 experimental groups

Randomized complete block design

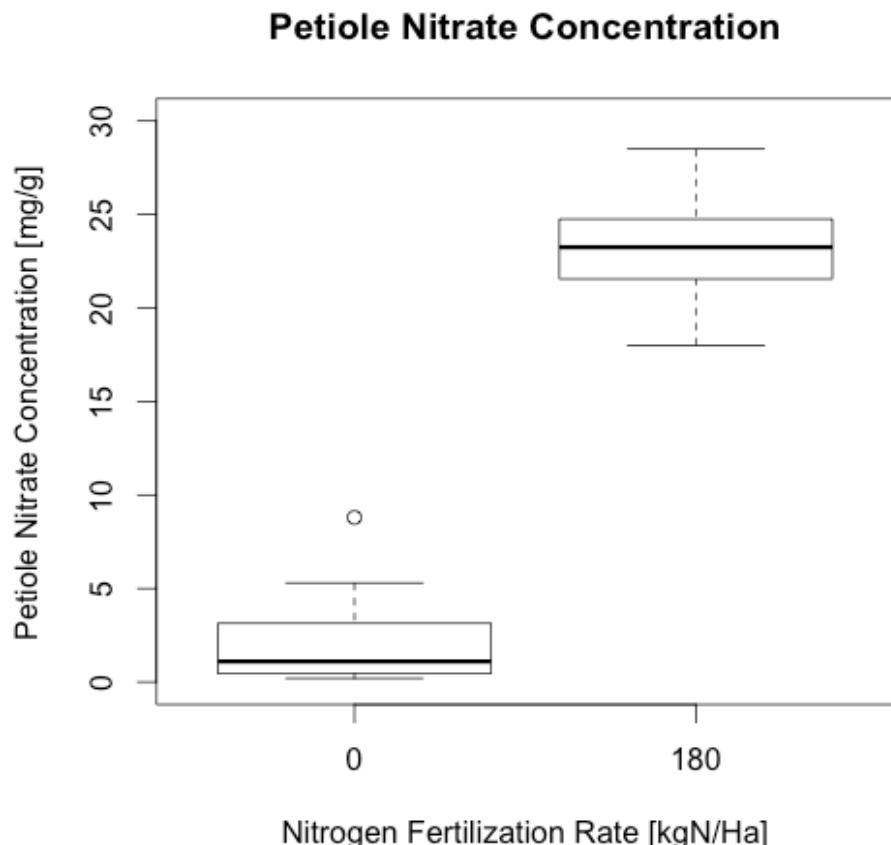
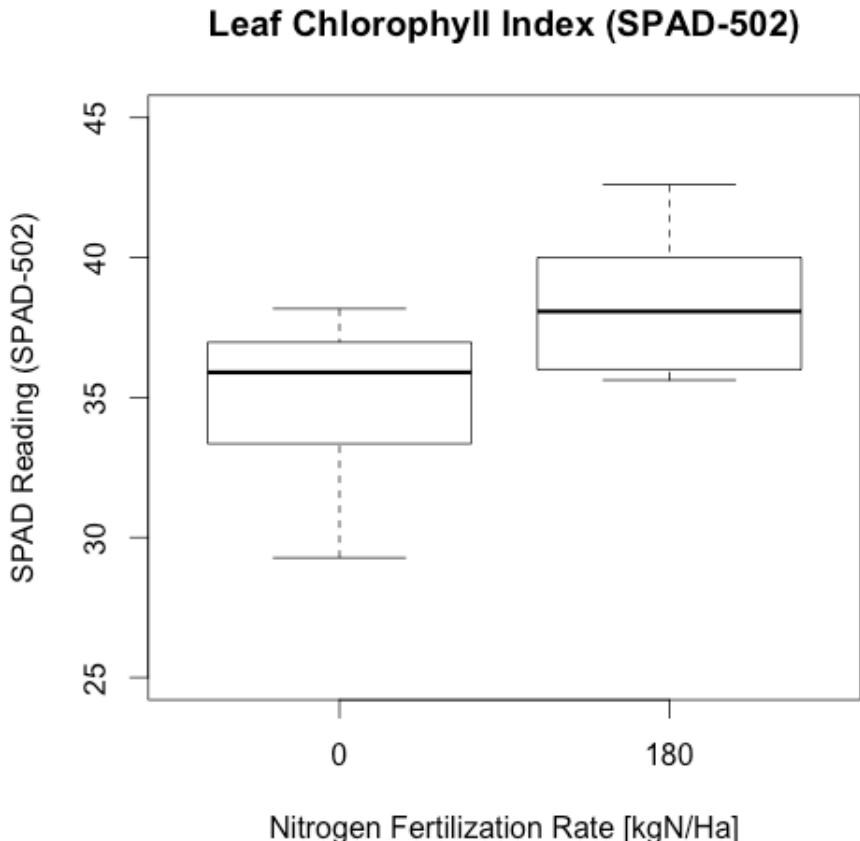
	Control (N-deficient) [0 kg N ha ⁻¹]		Treatment (added N) [180 kg N ha ⁻¹]	
	Time-point 1 July 25, 2012	Time-point 2 Aug. 8, 2012	Time-point 1 July 25, 2012	Time-point 2 Aug. 8, 2012
<i>S. tuberosum</i> cultivar				
Shepody	R1, R2, R3, R4	R1, R2, R3, R4	R1, R2, R3, R4	R1, R2, R3, R4
Russet-Burbank	R1, R2, R3, R4	R1, R2, R3, R4	R1, R2, R3, R4	R1, R2, R3, R4
Atlantic	R1, R2, R3, R4	R1, R2, R3, R4	R1, R2, R3, R4	R1, R2, R3, R4

Time-points: T1=before hilling, T2=after hilling

Biological Replicates: R1, R2, R3 and R4 consist of 20 plants each.

Sampled plants had different N status

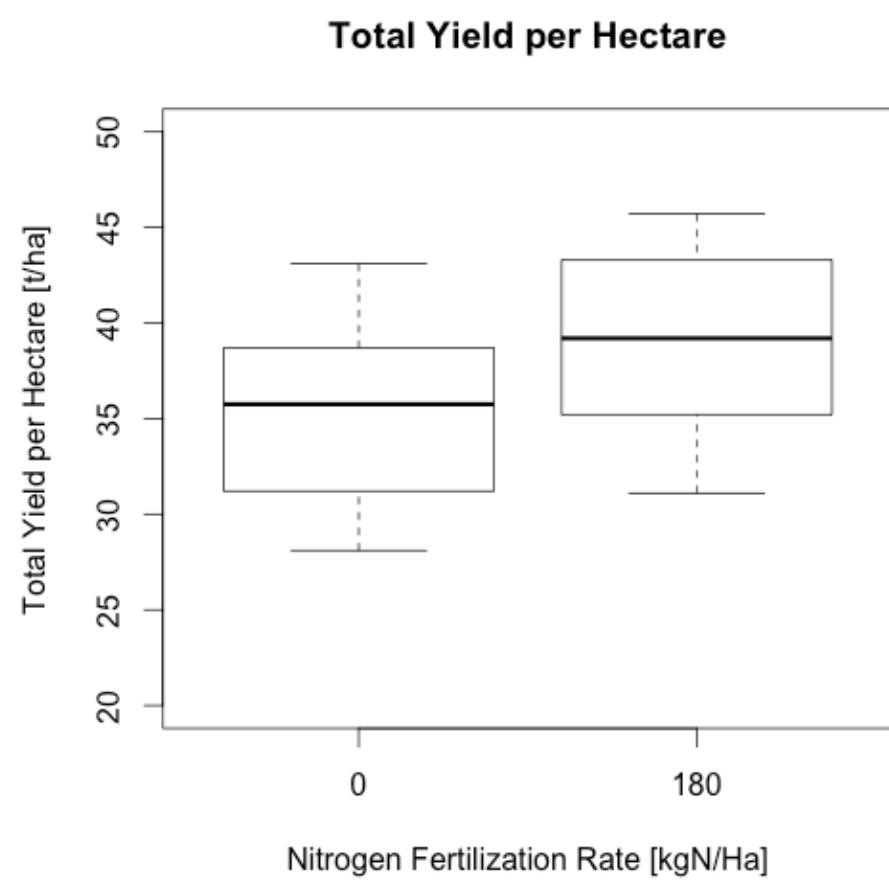
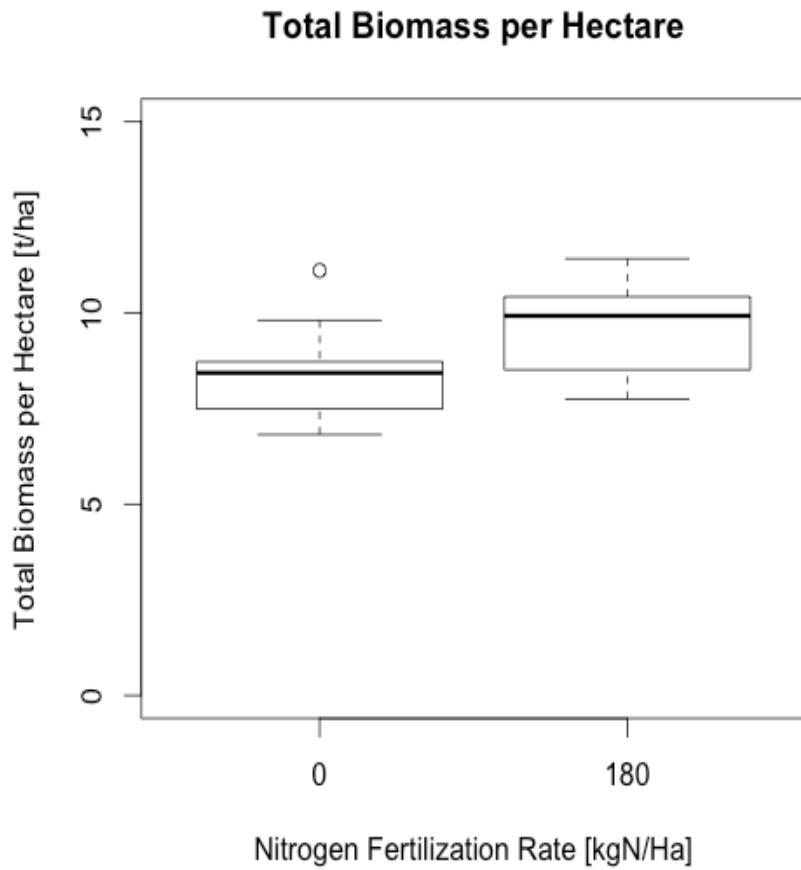
Phenotypic N sufficiency tests reveal significant differences between plants under different treatments:



For both tests: $df = 1$, $n = 48$, $\alpha = 0.05$

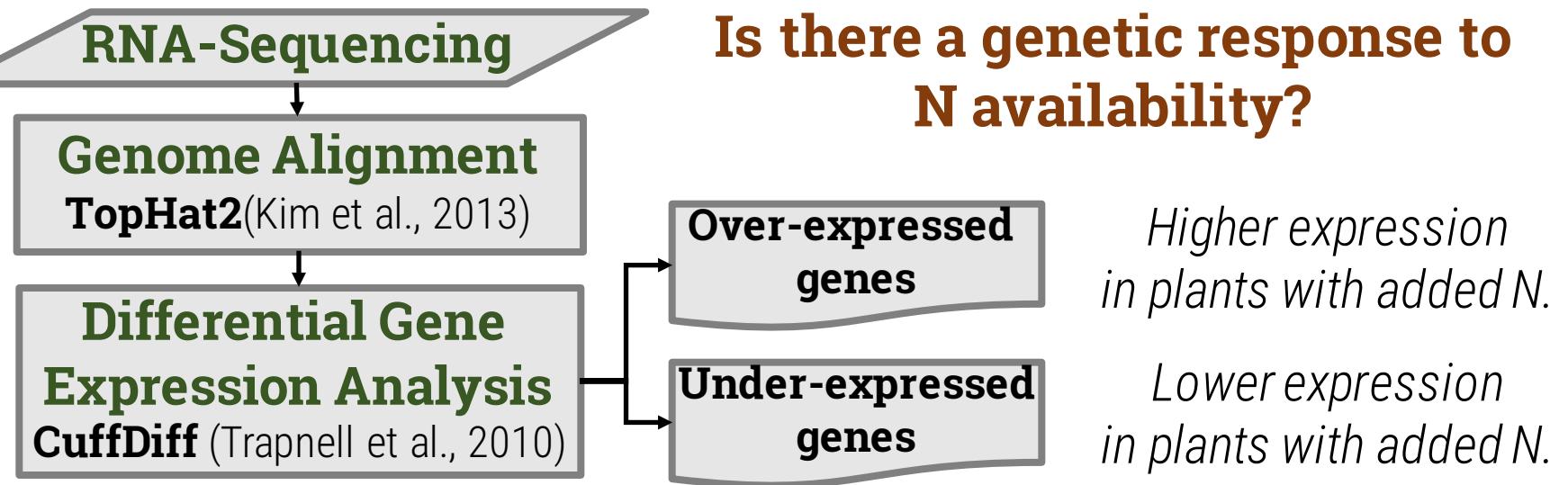
Importance of N fertilization

Impact of N fertilization on dry biomass and fresh yield:



For both tests: $df = 1$, $n = 24$, $\alpha = 0.075$

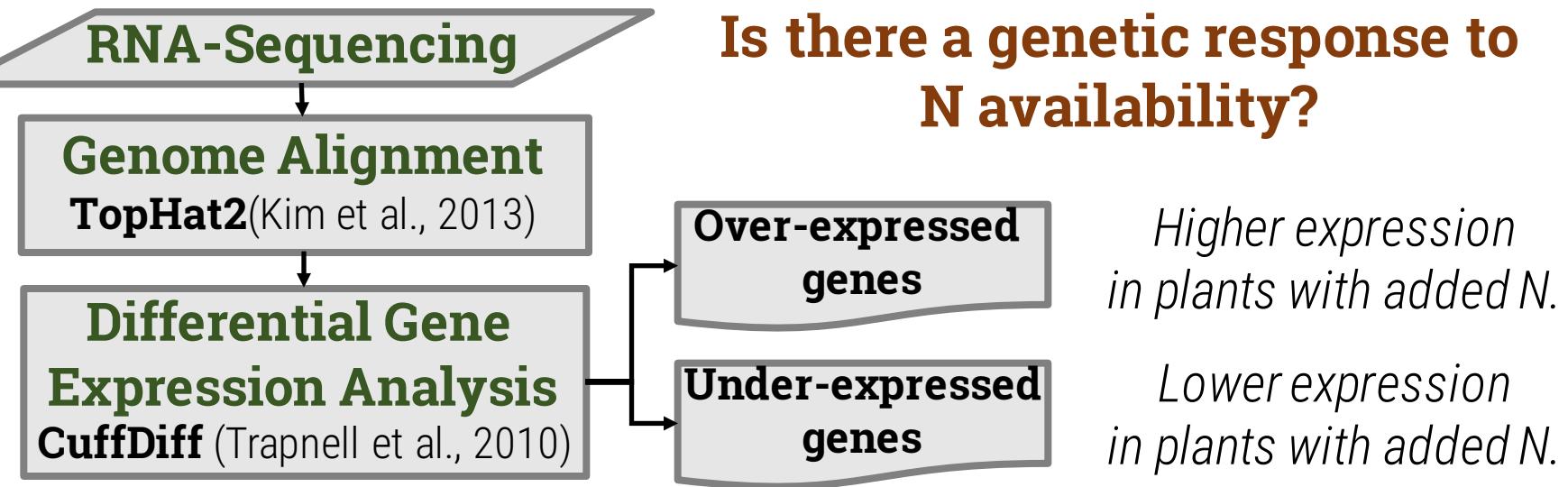
Obj. 1: Differentially expressed genes



Total number of differentially expressed genes found in each cultivar:

	Time-point 1		Time-point 2	
	Over	Under	Over	Under
Shepody	182	35	218	52
Russet-Burbank	64	47	116	18
Atlantic	393	33	149	40

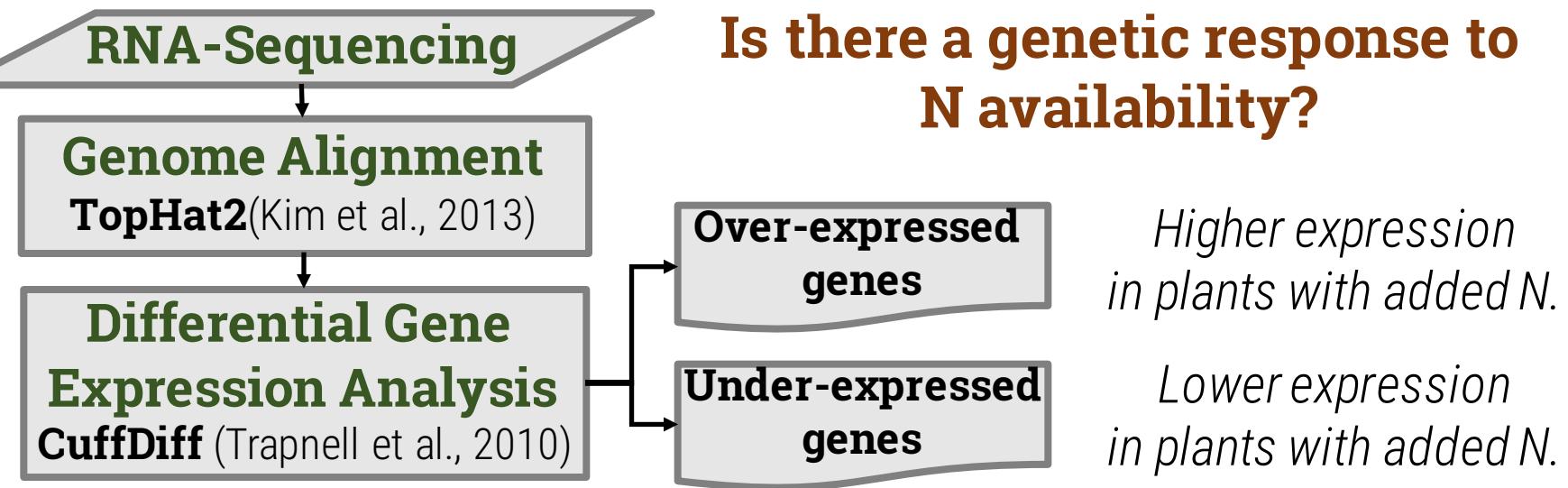
Obj. 1: Differentially expressed genes



Total number of differentially expressed genes found in each cultivar:

	Time-point 1		Time-point 2	
	Over	Under	Over	Under
Shepody				
Russet-Burbank	47	12	64	10
Atlantic				

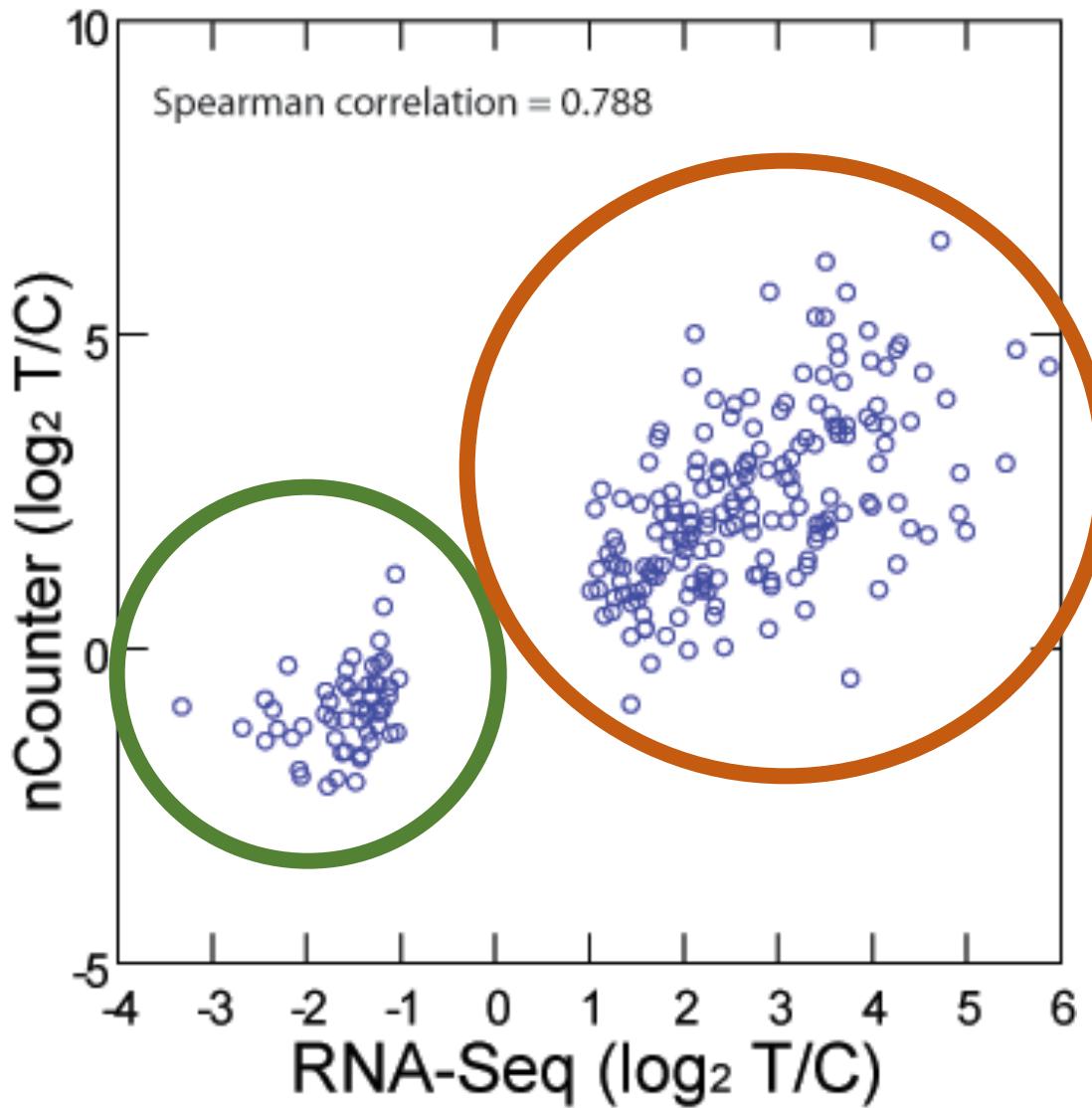
Obj. 1: Differentially expressed genes



Total number of differentially expressed genes found in each cultivar:

	Both Time-points	
	Over	Under
Shepody		
Russet-Burbank	30	9
Atlantic		

Obj. 1: Differentially expressed genes



nCounter Digital Analyzer

Correlation between \log_2 differences in gene expression measured by RNA-seq and nCounter.

Over-expressed genes
Under-expressed genes

Gene ID	Gene Name	Gene ID	Gene Name
Sotub08g007240	Cation transport regulator-like protein 2	Sotub08g014020	Chalcone isomerase
Sotub09g024290	Sulfate adenylyltransferase	Sotub11g007110	Plant-specific domain TIGR01615 family protein
Sotub06g008080	Male sterility 5 family protein (Fragment)	Sotub11g007090	Plant-specific domain TIGR01615 family protein
Sotub12g011100	Aminotransferase-like protein	Sotub10g018540	Aminotransferase like protein
Sotub03g018720	Alpha-glucosidase-like	Sotub04g026530	Peroxidase
Sotub12g027600	W.G.S. assembly reference scaffold set scaffold_4	Sotub04g023170	Unknown Protein
Sotub10g014450	Phenylcoumaran benzylic ether reductase 3	Sotub02g033060	NAD-dependent epimerase/dehydratase
Sotub10g024560	Glutathione S-transferase	Sotub01g049270	Tyrosine-protein kinase transforming protein Src
Sotub08g024220	Inositol 2-dehydrogenase like protein	Sotub11g012150	Amino acid transporter
Sotub05g024960	Amino acid transporter	Sotub01g005580	Glutamate decarboxylase
Sotub01g022620	Peptide methionine sulfoxide reductase msrB	Sotub07g009860	Peptide transporter
Sotub03g017290	Unknown Protein	Sotub02g036900	Cystine transporter Cystinosin
Sotub09g018850	Male sterility 5 family protein (Fragment)	Sotub04g027100	High affinity sulfate transporter 2
Sotub04g021910	Sulfate transporter	Sotub08g005390	Nitrate transporter
Sotub01g023000	Xylanase inhibitor	Sotub10g013960	High affinity sulfate transporter 2
Sotub12g012740	<i>Chloroplast lipocalin</i>	Sotub02g033320	Proline dehydrogenase
Sotub08g025870	Primary amine oxidase	Sotub09g009440	Cation/H ⁺ antiporter
Sotub09g023510	High affinity sulfate transporter 2	Sotub02g017430	Purine permease family protein
Sotub05g028860	Flowering locus T protein	Sotub12g031130	Poly(A) polymerase
Sotub01g049920	Nodule inception protein (Fragment)	Sotub09g010630	Hydrolase alpha/beta fold family protein
Sotub05g012720	Nodulin MtN21 family protein	Sotub09g029950	Cell wall protein
Sotub12g020880	Ubiquinone/menaquinone biosynthesis methyltransferase ubiE		

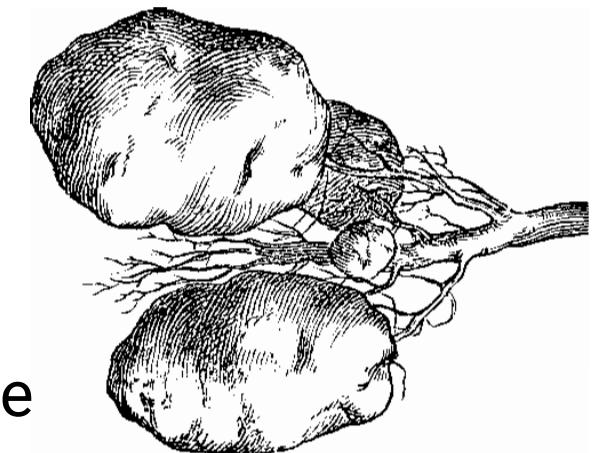
Obj. 1-a: Overrepresented pathways

► Amino acid metabolism

- Sotub10g018540 Aminotransferase like protein
- Sotub01g005580 Glutamate decarboxylase
- Sotub05g024960 Amino acid transporter
- Sotub11g012150 Amino acid transporter
- Sotub02g033320 Proline dehydrogenase

► Sulfate metabolism

- Sotub09g024290 Sulfate adenylyltransferase
- Sotub04g021910 Sulfate transporter
- Sotub04g027100 High affinity sulfate transporter 2
- Sotub10g013960 High affinity sulfate transporter 2
- Sotub09g023510 High affinity sulfate transporter 2



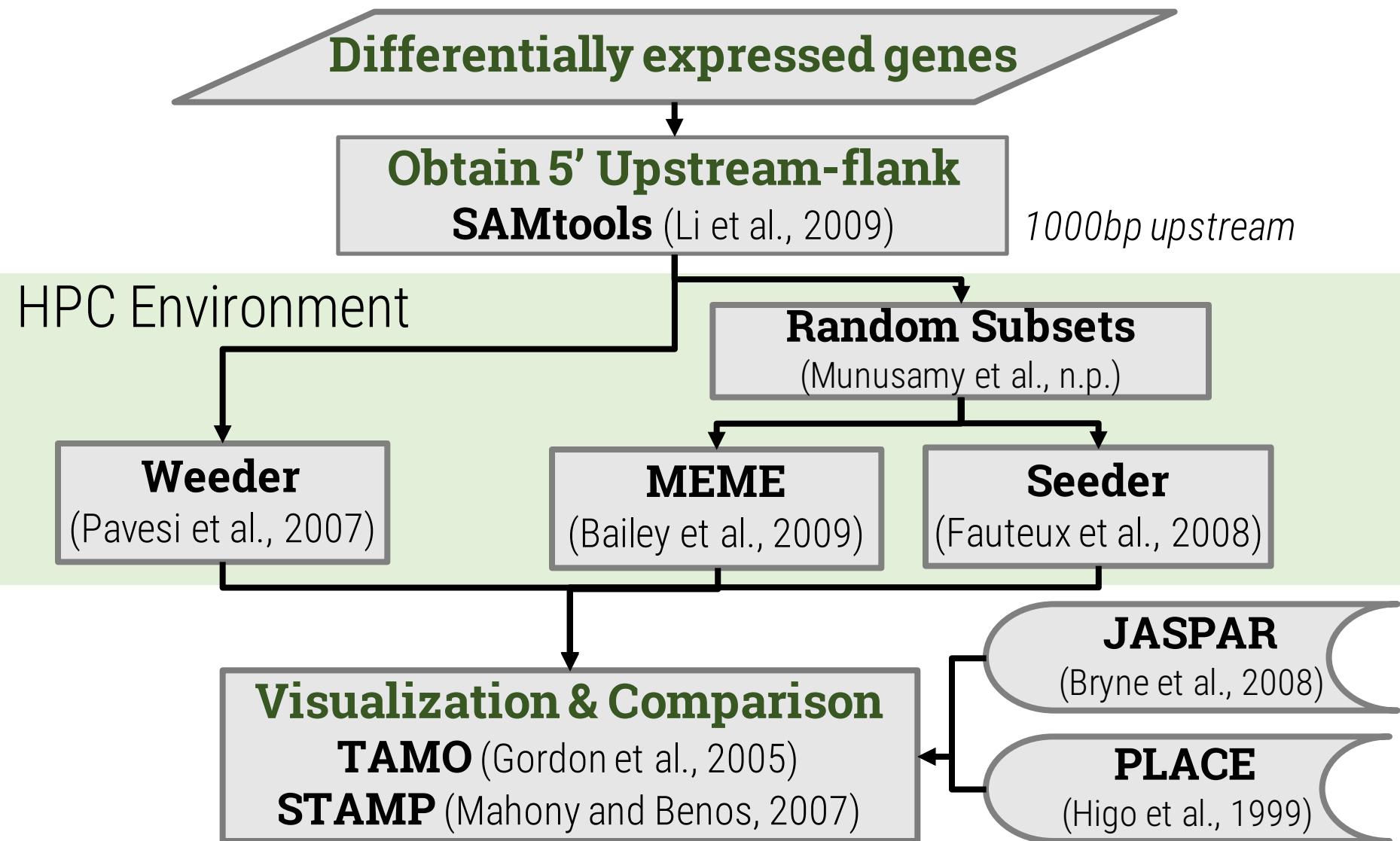
► Plant development genes

► Unknown function

Gene IDs based on ITAG1.0 Annotation
Underlined genes were under-expressed

Obj. 2: *De novo* motif discovery

What mechanisms regulate potato's N response?



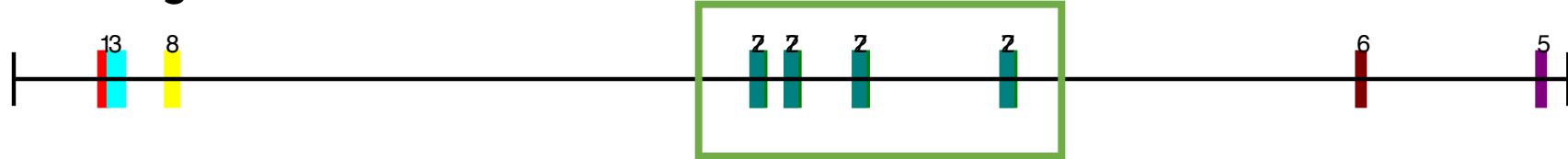
Obj. 2-b: Redundant motifs

Strategy 1: Clustering with *K-medoids*

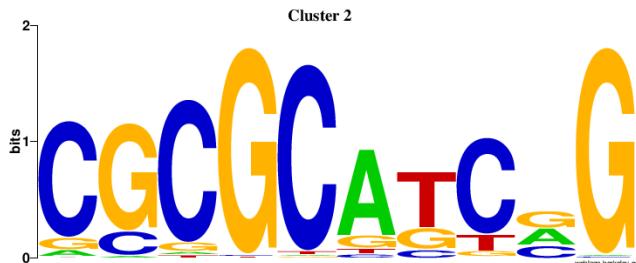
- **K-medoids**: iterative, combinatorial clustering method. TAMO implementation. (Gordon, et al., 2005)

Strategy 2: Mapping motifs within promoters

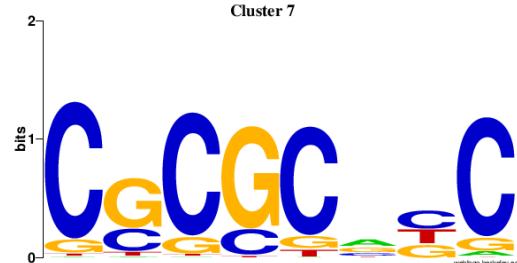
Sotub02g033060



Motif 1



Motif 2*



Both motifs only predicted by MEME

Obj. 2: *De novo* motif discovery

	Weblogo	Algorithm	PLACE	JASPAR
1		MEME	OCTAMERMOTIFTAH3H4 CNGATGCGCG -- GATCCGCG <i>histone; Oct; S-phase; CaMV 35S; NOS; meristem;</i>	MA0069.1_Pax6 -- CKGATGCGCG-- MANTSAGCGTGAA
2		Seeder	SITEIOSPCNA TGMACCTGGA -CCACCTGG- <i>PCNA; Site I; G-box; meristem;</i>	MA0086.1_sna TGMACCTGGA -- CACCTG--
3		Weeder	SURE2STPAT21 -ATACTC-- AATACTAAT <i>SURE; SURE 2; patatin; sucrose; tuber; root;</i>	MA0124.1_NKX3-1 -GAGTAT TAAGTAT
4		Seeder	OCTAMOTIF2 ATGCGG--- ATGCCGCGG <i>octamer; histone; meristem;</i>	MA0242.1_run_Bgb ATGCGG--- TTGCGGTTW
5		Seeder	3AF1BOXPSRBCS3 ANATAGAC----- AAATAGATAAATAAAAACATT <i>3AF1 box; promoter; AT-rich sequences; GATA; rbcS; rbcS-3; leaf; shoot;</i>	MA0011.1_br_Z2 GTCTATNT WNCTATTT

Obj. 2: *De novo* motif discovery

	Weblogo	Algorithm	PLACE	JASPAR
6		Seeder	TBOXATGAPB CTAAGT CAAAGT <i>GAPB; glyceraldehyde-3-phosphate dehydrogenase; light-activated transcription;</i>	MA0211.1_bap CTAAGT- TTAAGTG
7		Seeder	CATATGGMSAUR CATAGG CATATG <i>SAUR; NDE; auxin;</i>	MA0423.1_YER130C CATAGG--- NAWAGGGGN
8		Weeder	SB3NPABC1 -----TGTTCA--- AATTACTGTTCTAA <i>sclareol; ABC; transporter; SB3;</i>	MA0261.1_lin-14 TGAACA- -GAACRN
9*		Seeder	S1FBOXSORPS1L21 TACCAC TACCAT <i>S1F; S1F box; S1F-box; S1; plastid protein; RPS1; RPL21; leaf; negative;</i>	MA0002.1_RUNX1 -TACCAC---- NAACCACARWW

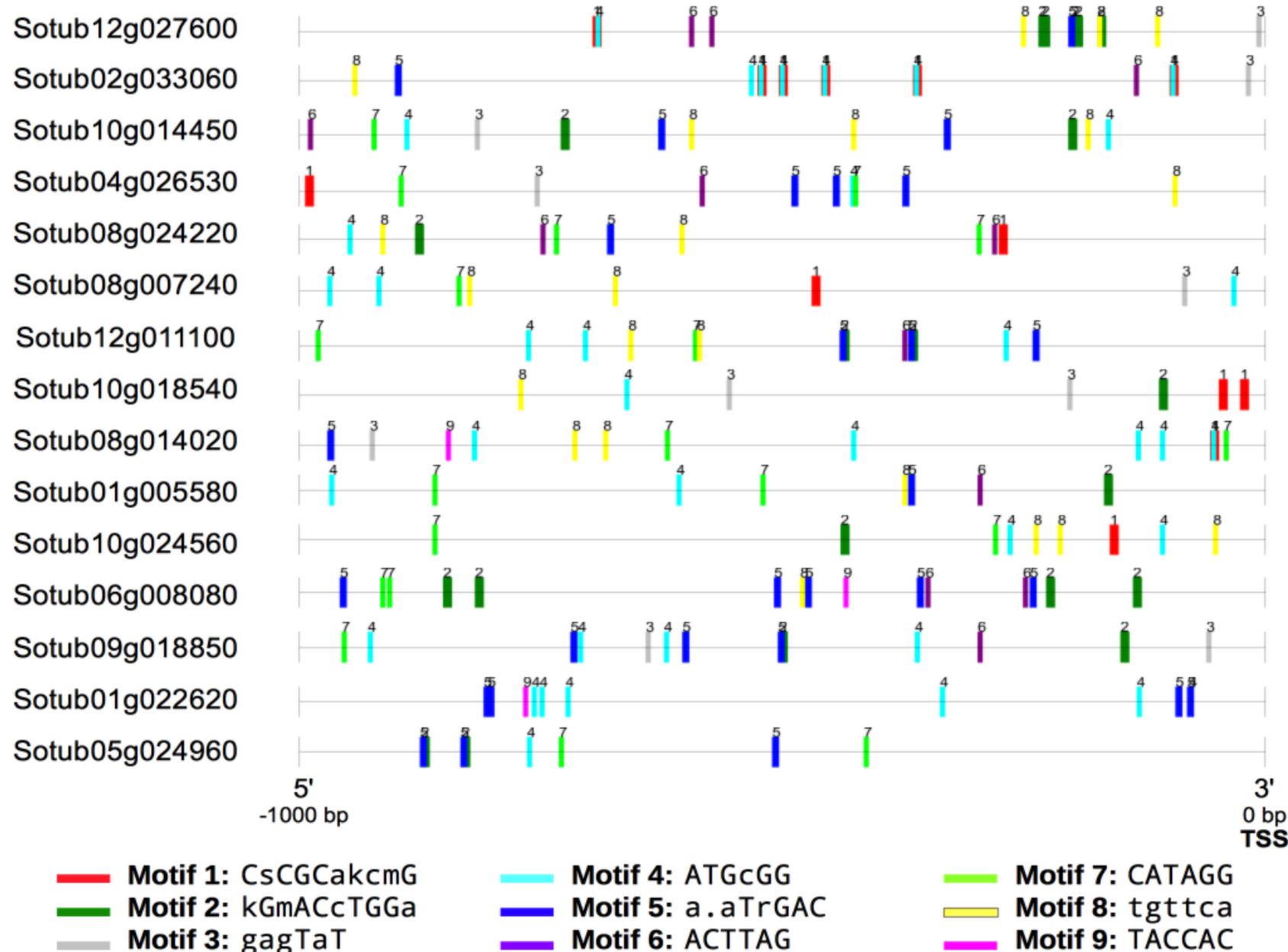
* Motif 9 was discovered in the upstream regions of *under-expressed* genes

Previously reported motifs

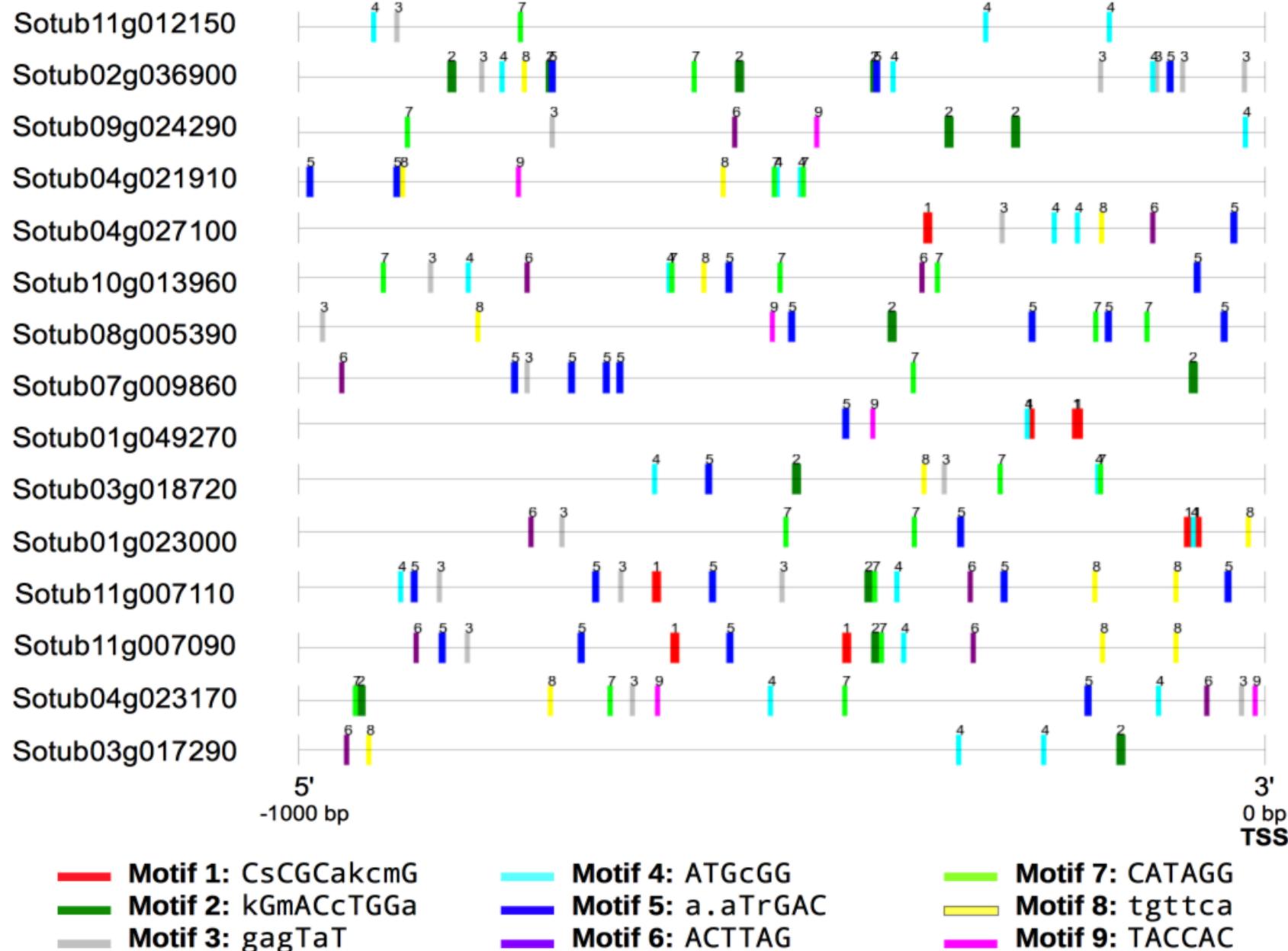
Upstream flanking regions of N responsive genes were searched for instances of previously reported N regulatory motifs:

- Nitrate Related *cis*-Element (NRE)
 - Not found.
 - Found in the upstream region of 2 out of the 3 potato NIR genes.
- Maize nitrogen-related motifs
 - A few instances found, however not significantly overrepresented.

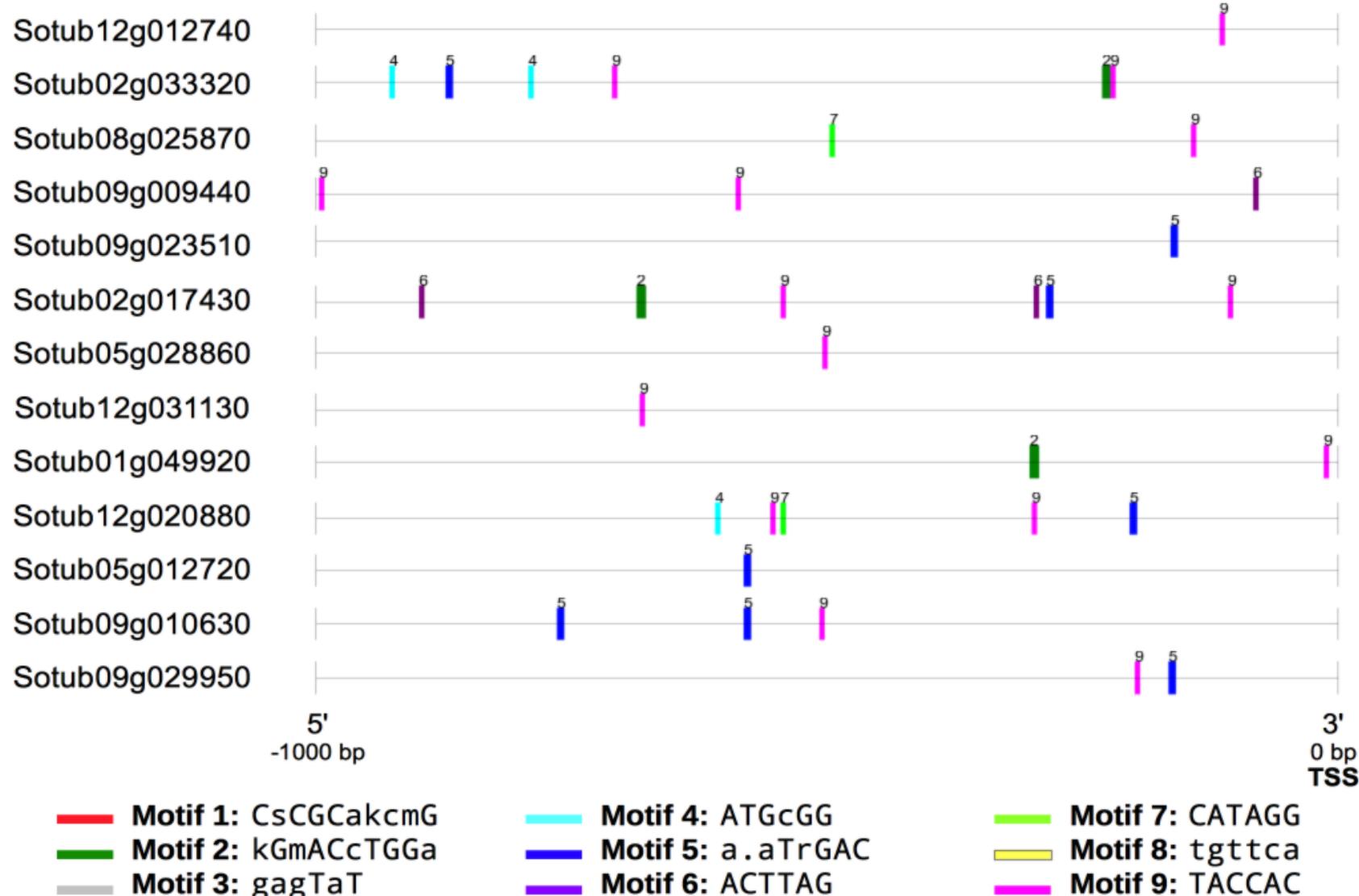
Upstream region of over-expressed genes



Upstream region of over-expressed genes



Upstream region of under-expressed genes*



* Includes upstream region of genes found to be under-expressed in only one time-point

Conclusions

- Identified 39 N responsive genes across three cultivars and two developmental time-points.
- Predicted and mapped 9 putative regulatory motifs in the 5'-upstream region of N responsive genes.
- Bioinformatics tools can identify potential regulatory mechanisms which can later be experimentally validated.

www.nature.com/scientificreports/

SCIENTIFIC REPORTS



OPEN

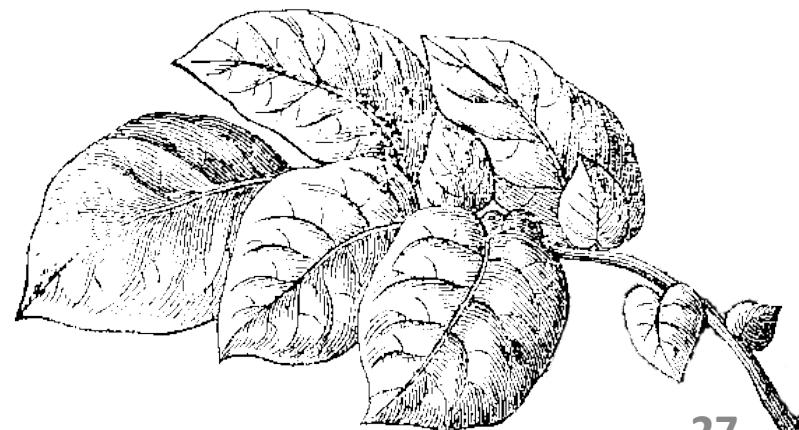
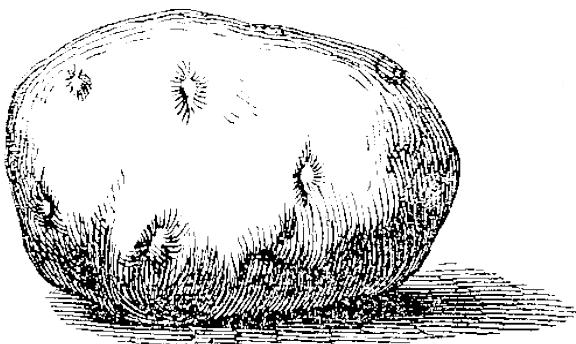
Received: 23 December 2015
Accepted: 25 April 2016
Published: xx xx xxxx

The nitrogen responsive transcriptome in potato (*Solanum tuberosum* L.) reveals significant gene regulatory motifs

José Héctor Gálvez¹, Helen H. Tai², Martin Lague², Bernie J. ZebARTH² & Martina V. Strömvik¹

Future work

- **Experimental validation of putative regulatory motifs** found in N responsive genes.
 - Identify regulatory elements binding to motifs.
 - Verify potential effects on phenotype.
- **Identify cultivar-specific gene regulation mechanisms** in potato.
 - Requires cultivar-specific genomic information.



Acknowledgments

Supervising Committee

Martina Strömvik, Ph.D. (Supervisor)

Helen Tai, Ph.D.

Jean-Benoit Charron, Ph.D.

Agriculture and Agri-Food Canada

Martin Lagüe, Ph.D.

Bernie ZebARTH, Ph.D.

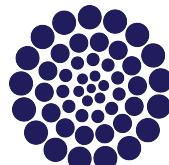
McGill University Department of Plant Science Graduate Excellence Award



McGill



Agriculture and
Agri-Food Canada



CONACYT
Consejo Nacional de Ciencia y Tecnología

Agriculture et
Agroalimentaire Canada

Strömvik Laboratory

Eva Charbonneau-Bérubé

Maria Kyriakidou

