Regulus infers signed, context-dependent and process-based regulatory circuits between few cell types

Marine LOUARN[1,2,4], Fabrice CHATONNET[2,3], Guillaume COLLET[1], Ève BARRÉ[1], Thierry FEST[2,3], Anne SIEGEL[1] & Olivier DAMERON[1]

1 - Univ Rennes, CNRS, Inria, IRISA - UMR 6074, F-35000 Rennes, FRANCE
2 - UMR_S 1236, Université Rennes 1, INSERM, Etablissement Français du Sang, F-35000 Rennes, FRANCE
3 - Laboratoire d'Hématologie, Pôle de Biologie, CHU de Rennes, F-35033 Rennes, FRANCE
4 - Current address: Pavlidis Lab, University of British Columbia, Vancouver, BC, CANADA

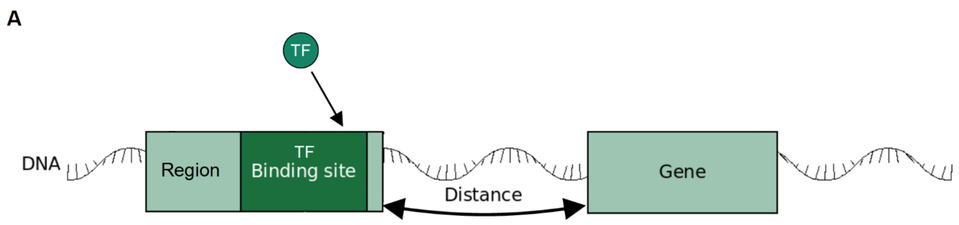
Motivation:

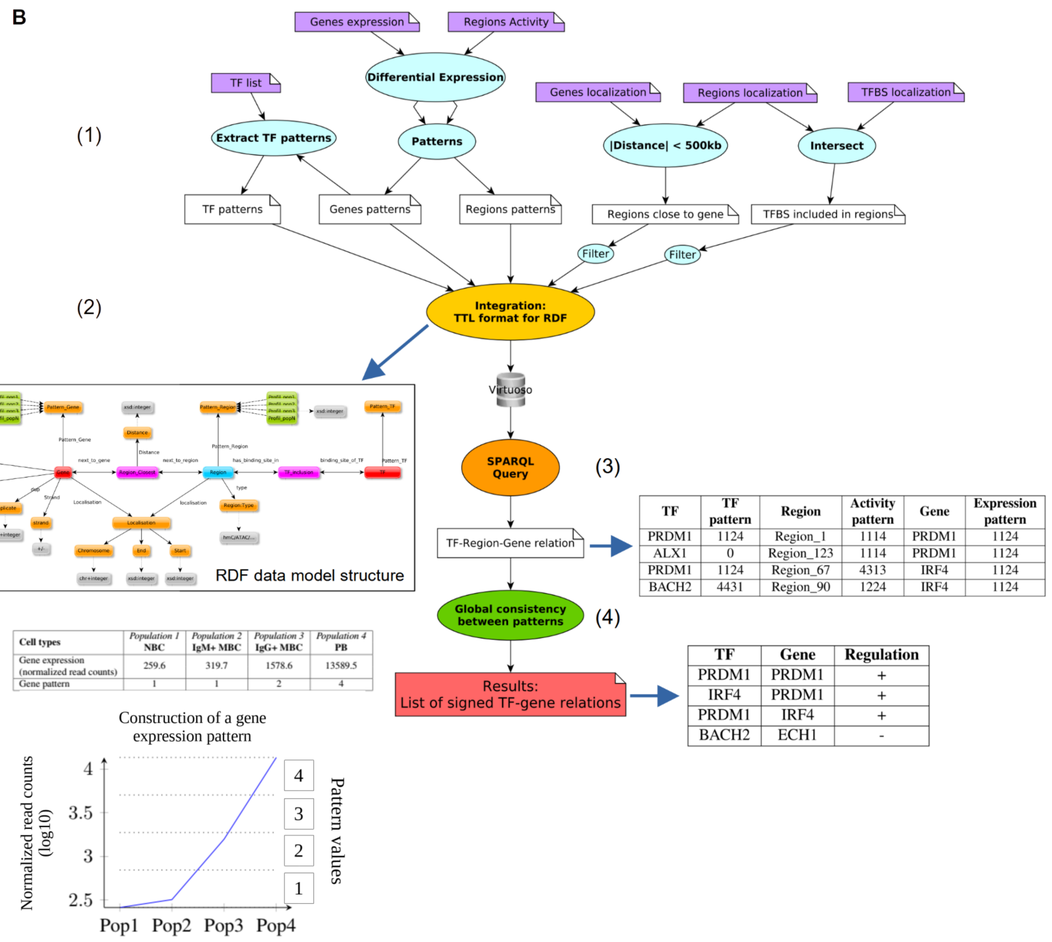
• Understanding the dynamic system & the keys regulators of cell differention between closely related cell types, which are finely tuned by regulatory mechanisms.
• Most existing methods for inferring regulation are not applicable to common experimental or clinical settings, where the number of samples is limited.
• We created a context specific regulatory circuits inference tool, based on our previous work with Semantic Web technologies (*Louarn et al., 2019, 2022*).

Results:

• Our context-specific regulatory inference tool, *Regulus* computes TF-gene relations from gene expressions, regulatory region activities & TF binding sites data.
• Expression & activity patterns describe data dynamics.
• Data are integrated & queried to retrieve all potential TF-region relations.
• Relations filtered & signed using global consistency constraints on patterns.
• *Regulus* is well suited for scarce-sample & closely-related cell populations settings.

Regulus pipeline:





TF	TF pattern	Region	Activity pattern	Gene	Expression pattern
PRDM1	1124	Region_1	1114	PRDM1	1124
ALX1	0	Region_123	1114	PRDM1	1124
PRDM1	1124	Region_67	4313	IRF4	1124
BACH2	4431	Region_90	1224	IRF4	1124

TF	Gene	Regulation
PRDM1	PRDM1	+
IRF4	PRDM1	+
PRDM1	IRF4	+
BACH2	ECH1	-

C

Cell types	*Population 1* NBC	*Population 2* IgM+ MBC	*Population 3* IgG+ MBC	*Population 4* PB
Gene expression (normalized read counts)	259.6	319.7	1578.6	13589.5
Gene pattern	1	1	2	4

Construction of a gene expression pattern

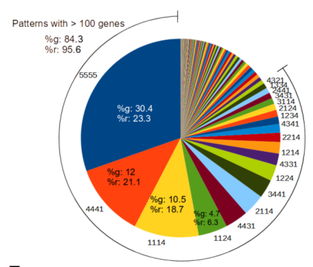
A - Biological relationships between the entities used to compute the regulatory circuit.
B - Different steps of the pipeline.
(1) Pre-process to explicit relations between entities & activity patterns creation
(2) RDF graph to formalize entities relations & generate a data model structure
(3) SPARQL query to extract all TF-Region-Gene relations & respective patterns
(4) Logical consistency constraints to get a signed & filtered network of unique TF-gene relations
C - Gene expression pattern construction.

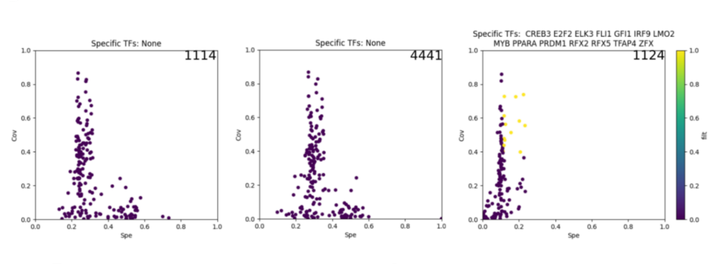
Application to B cells & Further filtering

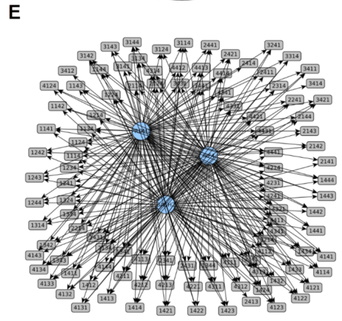
Cell types	*Population 1* NBC	*Population 2* IgM+ MBC	*Population 3* IgG+ MBC	*Population 4* PB
PRDM1 gene expression (normalized read counts)	259.6	319.7	1578.6	13589.5
PRDM1 gene pattern	1	1	2	4

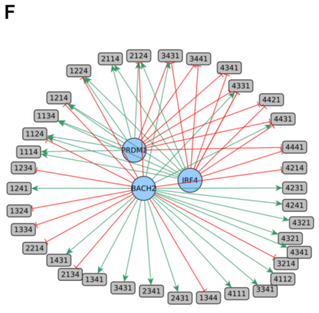
B

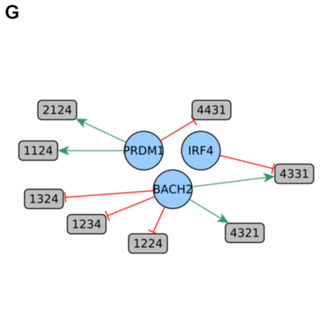
	Class	Number of entities
Input	Gene	26,802
	ATAC_region	58,848
	ATAC_close_gene	496,282
	TF	594
	TFBS	3,460,101
	Patterns	109
Result	TF-regions-genes relations	5,635,099
	TF-Gene relations	314,965











A - Known biology behind B cell differentiation: with PRDM1, IRF4 & BACH2 known regulators.
B - Main in/outputs of Regulus for B-cell differentiation regulatory network.
C - Number of genes per expression patterns: 18 patterns contain >100 genes.
D - Coverage & specificity distribution for all TFs that target specific patterns. Yellow dots indicate TFs over threshold.
E to G - Interaction graphs of 3 known regulators & their targeted patterns
(E) before filtering
(F) after filtering with consistency step
(G) after filtering with coverage & specificity. Note that relations are consistent with the known roles of BACH2, IRF4 and PRDM1 during the PB differentiation.

Applied to B cell differentiation data, Regulus identifies :
▶ known regulators: IRF4, PRDM1, BACH2 and PAX5
▶ 6 candidate new regulators (FOXJ3, KLF16, TFAP4, TGIF1, ZNF219, ZNF75A)

Conclusion

• *Regulus* provides signed, globally consistent TF-gene relations over few samples of closely-related cell types
• Applied to B cells, *Regulus* identifies both known & potential new regulators

Acknowledgment

We thank the GenOuest bioinformatics core facility for providing the computing infrastructure. (https://www.genouest.org)