



PRECURSOR

SCIENTIFIC
NETWORK, then
EXPLORATORY
PROJECT

2024-2028

Coordination

Marie-Laure Martin

IPS2 JRU

marie-laure.martin@inrae.fr

Sylvie Coursol

IJPB JRU

sylvie.coursol@inrae.fr

Key words

Plants

Cis-regulatory elements

Transposable elements

Knowledge base

Artificial Intelligence

INRAE participating units

[IPS2](#)

[IJPB](#)

[URGI](#)

[MIA Paris-Saclay](#)

Partnerships

[CIRAD \(AGAP\)](#)

[IRD](#)

[Université Clermont Auvergne](#)

Expanding our fundamental knowledge of gene -proximal regions to improve selection models

Gene transcription is an essential process in the adaptive response of plants to environmental constraints. The interdisciplinary scientific consortium PRECURSOR aims to investigate and better understand how this process takes place in the proximal regions of genes to ultimately improve the predictive power of selection models.

Context and challenges

Transcription, the first stage of gene expression and protein synthesis, is tightly regulated by a number of molecular elements. *Cis*-regulatory elements, which consist of short DNA sequences, regulate gene expression via *trans*-acting factors that bind to the *cis*-regulatory elements.

Modifying gene expression through regulators

Cis-regulatory sequences are present in high density in the proximal regions of genes, but their characterization, an essential prerequisite for their use, remains incomplete. Recent projects have mapped DNA sequences preferentially located (known as PLMs) in these regions (in *Arabidopsis thaliana* and maize), with nearly 80% still unassigned in databases, although some are supported by MNase-defined cisome Occupancy Analysis sequencing. Additionally, numerous studies have shown that transposable elements (TEs) can include *cis*-regulatory sequences. When TEs are inserted near a gene, they can then affect the transcription of neighboring genes by recruiting additional *trans* factors.

These two data sources (PLMs and TEs) are promising as they allow for the large-scale characterization of potential *cis*-regulatory elements. However, to gain a true understanding of proximal regions, these structural data need to be coupled with expression data. Original approaches using artificial intelligence may offer a promising way to integrate these biological data, thereby enabling the prediction of key genes and their regulatory networks.

However, there are few opportunities for teams of experts working in these areas to come together with their different and complementary skills. The PRECURSOR scientific network was therefore established to overcome this obstacle by creating an interdisciplinary network of experts to address this topic.





© Freepik

Goals

PRECURSOR brings together scientific teams working at the interface between biology (molecular science, genetics, physiology) and formal science (statistics, computer science, bioinformatics), to investigate different species (maize, wheat, sorghum) and gain a consolidated vision of the genetic basis for traits of agronomic interest that encompasses both structural and expression data.

The aim is to collaboratively advance the mapping and predictive power of *cis*-regulatory elements in the proximal regions of genes, taking into account the overall complexity of the question and the complementarities/differences between the species studied.

PRECURSOR's main objective is to form an interdisciplinary scientific consortium based on the unprecedented integration of heterogeneous data to gain a better understanding of the proximal regions of genes and ultimately to develop new alleles of agronomic interest and improve the predictive power of selection models.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertise
Plant Biology and Breeding	<u>IPS2</u>	Bioinformatics of <i>cis</i> -regulatory elements, statistics of omics data
	<u>IJPB</u>	Biology of <i>cis</i> -regulatory elements; maize, environmental constraints, digestibility, functional genomics
	<u>URGI</u>	Information technology, knowledge bases, transposable elements
Mathematics and Digital Technologies	<u>MIA Paris-Saclay</u>	Artificial Intelligence methods
External partners		Expertise
CIRAD	<u>AGAP (PRECURSOR-2 follow-up exploratory project)</u>	Quantitative genetics, sorghum, functional genomics
IRD	<u>DIADÉ</u>	Biology, tropical cereals, root systems
University of Clermont Auvergne	<u>GDEC</u>	Molecular physiology of responses to biotic and abiotic stress, wheat, fungal pathogens, water stress

Outcomes

Metaprogramme
DIGIT-BIO



Contact: digitbio@inrae.fr

<https://digitbio.hub.inrae.fr/>

Building an interdisciplinary community

The consortium, created in 2024, brings together 12 members with complementary profiles (biologists, geneticists, experts in bioinformatics and informatics, and statisticians) across three sites (Ile-de-France, Clermont-Ferrand and Montpellier). It enables members to work together effectively and to share tools. These interactions have encouraged the emergence of a shared language and an improved mutual understanding of discipline-specific constraints and benefits.

Exploration and critical analysis of existing approaches to the detection, characterization and prediction of Cis-regulating elements (CREs)

This collective work enabled the relevance of different methodological approaches to be assessed, data needs to be identified, and the necessary conditions for informed biological modelling to be defined. Initial tests, which involved, in particular, the mapping of CREs in various plant species, provided the evidence base for this assessment.

The bibliographical analysis conducted brought the limitations of current approaches into focus, particularly their lack of biological interpretability, their limited transferability between species, and the difficulty they have in predicting complex phenotypes. It helped to build a clearer picture of the scientific obstacles that still lie in the path of the effective application of such methods to the improvement of plants. This work enabled an opinion article to be produced, which is currently in press. It proposes a shift in the study of CREs toward approaches that explicitly integrate transcriptional regulatory mechanisms, along with their organization and evolutionary context, in order to move beyond a purely statistical predictive framework and strengthen model interpretability.

Future directions: launch of the Precursor-2 Exploratory Project (2026-2028)

The PRECURSOR network has leveraged the development of new approaches, not least the PRECURSOR-2 exploratory project, which successfully obtained funding through the 2026 DIGIT-BIO AMI. PRECURSOR-2 will make use of recent artificial intelligence methods, including deep learning models and language models applied to DNA sequences, seeking to identify and systematically map CREs. By applying such approaches to multiple genomes (both nuclear and plastidial) and to species, the project intends to assess their robustness and transferability. A second major line of inquiry for PRECURSOR-2 is the exploration of the functional associations between CREs and transposable elements, in which sequencing data, functional data and adapted modelling frameworks are combined. Last, the project also aims to test the explanatory and predictive capabilities of CREs with regard to gene expression and traits of agronomic interest, particularly under conditions of environmental stress. The investigatory principle underlying all these actions is the integration of CREs into quantitative genetics and systems genetics approaches, thereby complementing traditional approaches based on anonymous markers.

