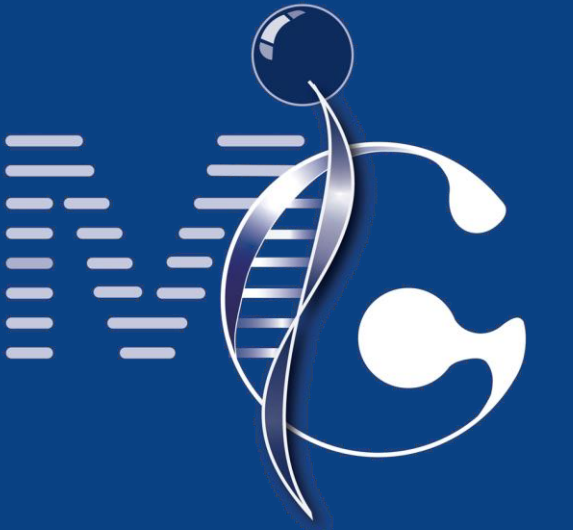


# StableMate: a new statistical method to select stable predictors in omics data

A case study of characterising cell identity transition of glioblastoma tumour-infiltrating microglia with scRNA-seq data



Yidi Deng<sup>1,2</sup>, Jiadong Mao<sup>1</sup>, Jarny Choi<sup>2,†,&</sup>, Kim-Anh Lê Cao<sup>1,\*,&</sup>

<sup>1</sup> Melbourne Integrative Genomics, School of Mathematics and Statistics, The University of Melbourne, Australia  
<sup>2</sup> Department of Anatomy and Physiology, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne & Bioinformatics and Cellular Genomics, St Vincent's Institute, Australia

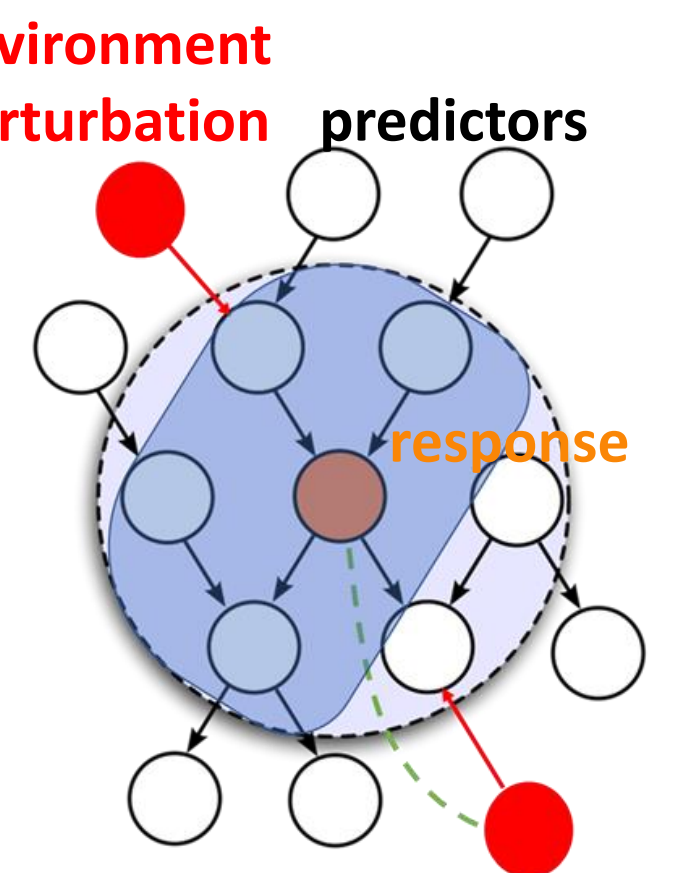
## Highlight

- We developed **StableMate**<sup>1</sup>, a statistical approach which identifies statistical associations that are environment-specific or –agnostic and stable.
- Environment refers to a biological condition or an experiment from which samples are collected.
- We applied StableMate to a scRNA-seq dataset from Darmanis et al. (2017), who sequenced myeloid cells from the core and the periphery of Glioblastoma (GBM).
- We first characterized cell trajectory between the two locations, core and periphery.
- Our StableMate analysis revealed transcriptional activities of microglia that are consistently seen or specific to the two locations during its transition to tumour associated microglia.

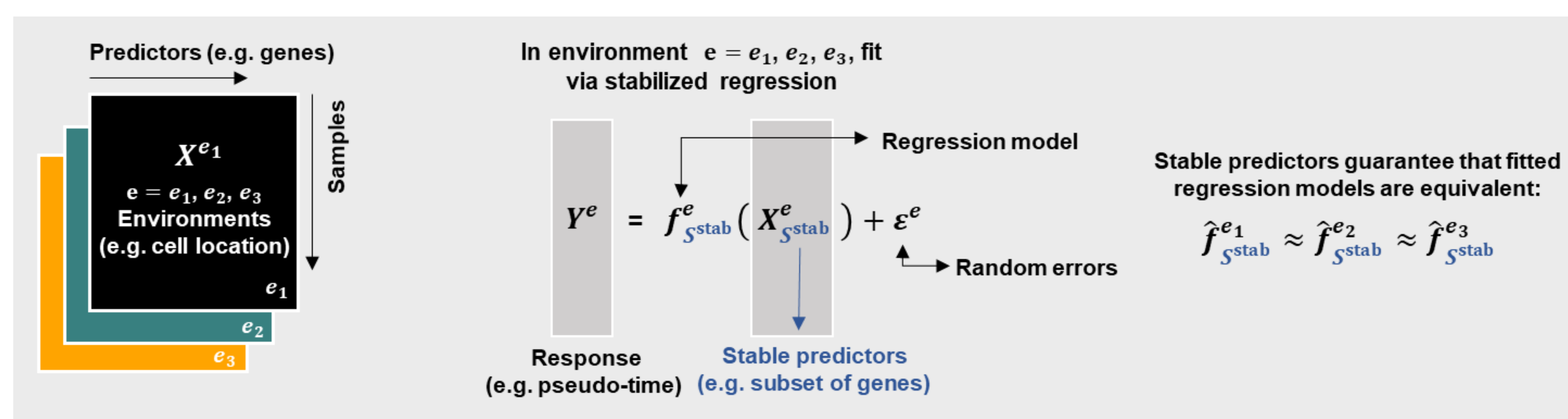
## Link between Stability & Causality

Stable associations are approximately causal

(In a regression setting, conditioning on stable predictors, response and environment perturbation are independent)



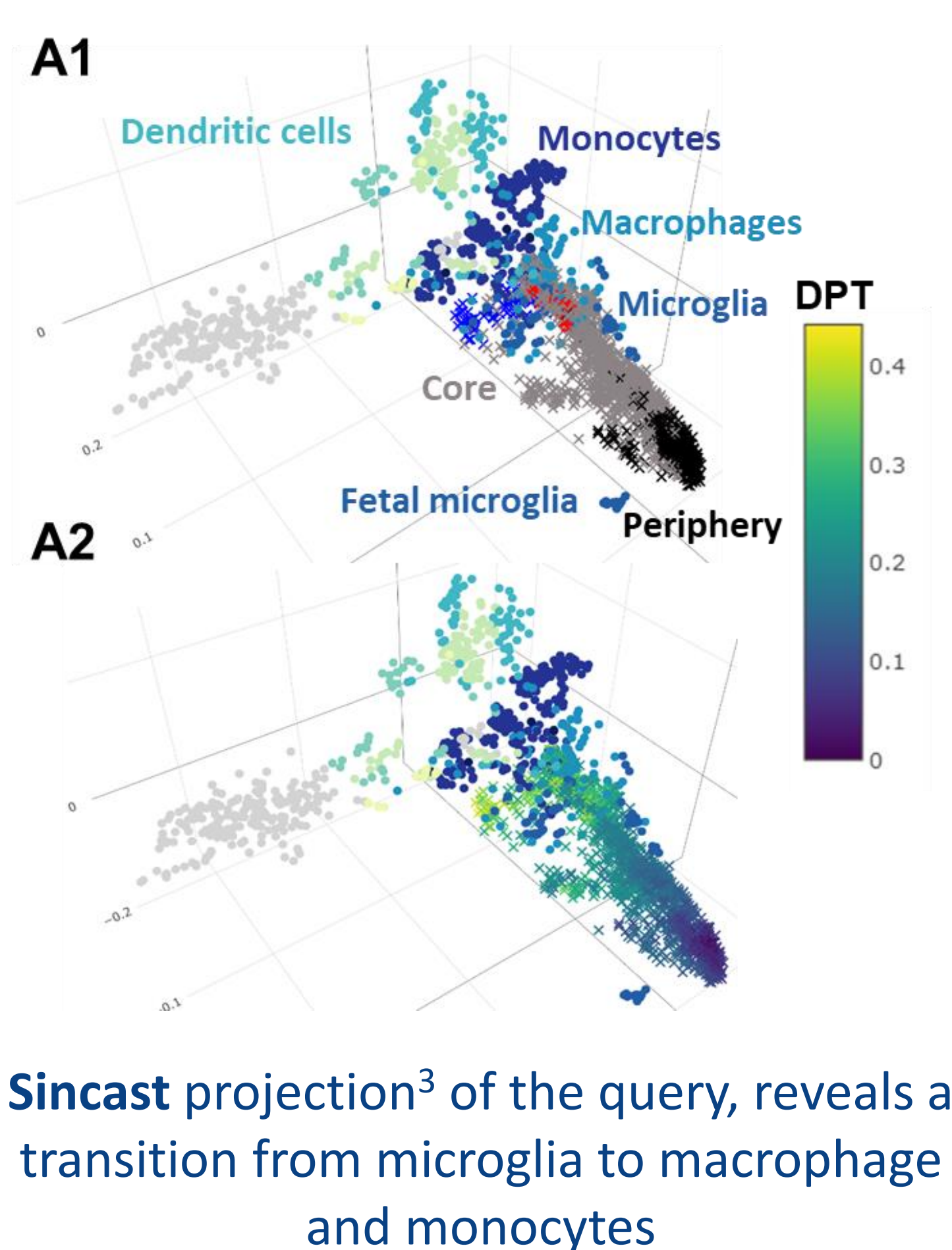
## Method



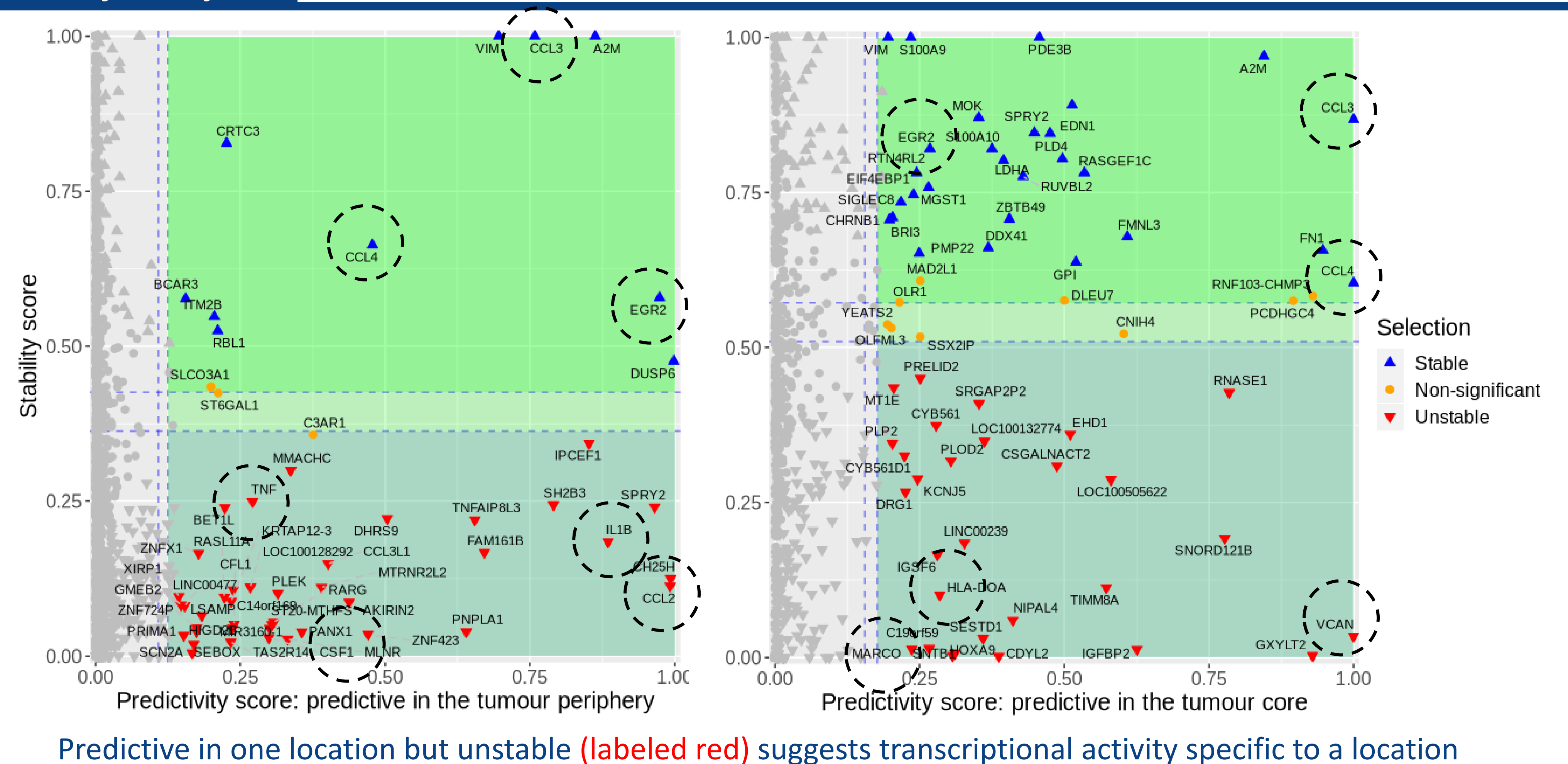
Stable associations: functional dependencies that are consistent across cell locations

- We improved a **stochastic selector**: stochastic stepwise (ST<sup>2</sup>) variable selection to select genes.
- Run ST<sup>2</sup> for K times to select **highly predictive genes**.
- Within each selected predictive gene set, run ST<sup>2</sup> for K times to select for **stable genes**.
- Calculate **importance scores** of predictive genes based on weighted average of selection.
- Build a regression ensemble based on the selection ensemble

## Data



## Stability analysis

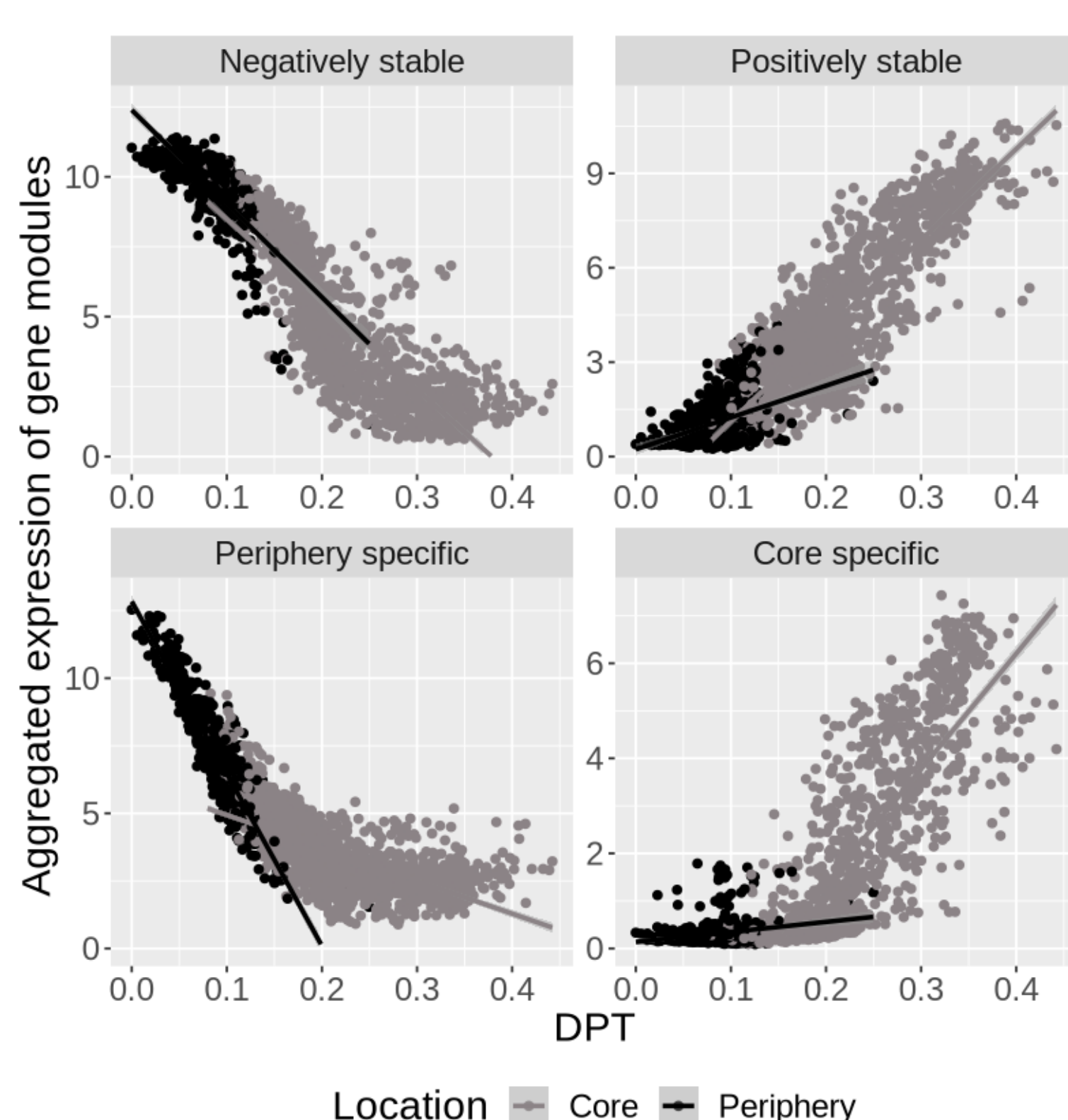


## Major patterns

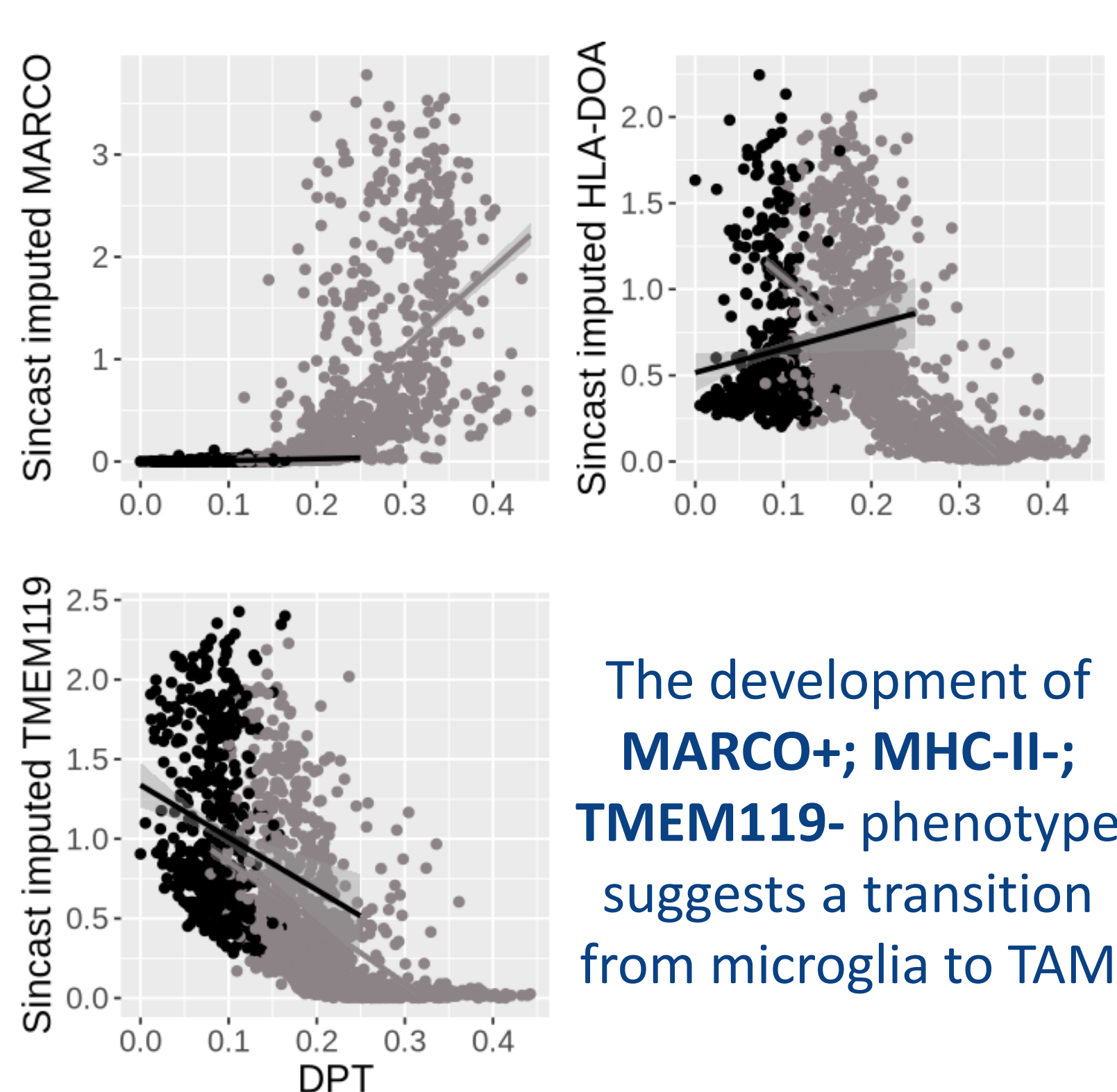
**Negatively Stable:**  
CCL3, CCL4, EGR2, CD83  
signify pre-activation

**Periphery specific:**  
TNF, CCL2, CSF1,  
IL1B signify  
polarisation

**Core specific:**  
VCAN, GXYLT2  
signify the  
development of TME



## Core-specific M2 TAM markers



## Conclusion

- StableMate make biological insights that cannot be done by traditional statistical analysis (e.g., co-expression analysis).
- Periphery myeloid cells were primarily microglia undergoing polarisation.
- Pro-inflammatory microglia seem to transiently exist at the start of the core-stage of the transition and were rapidly reprogrammed into M2-like TAM.
- Microglia pre-activation markers seem to active consistently in both locations regardless of the pressure from tumour microenvironment.
- StableMate can be applied to wide range of biological data types and questions as demonstrated in our two other case studies.

## References

- Deng et al. 2023 "StableMate: a new statistical method to select stable predictors in omics data." bioRxiv
- Deng et al. 2022 "Sincast: a computational framework to predict cell identities in single-cell transcriptomes using bulk atlases as references." Briefings in Bioinformatics 23.3
- Xin et al. 2012 "Stochastic stepwise ensembles for variable selection." J oComp and Graphical Stat 21.2

E: yidid@student.unimelb.edu.au  
Github: <https://github.com/meiosis97>

## Funding

ARC DP200102903, NHMRC Career Development fellowship GNT1159458 (KALC) and the Melbourne Research Scholarship (YD).