



# Geographically weighted methods: PCA

Dr SIMON J COCKELL  
ELEFTHERIOS (LEFTERIS) ZORMPAS

Biosciences Institute,  
Faculty of Medical Sciences,  
Newcastle University  
23/07/2023

© ICBAM research group, Newcastle University, UK

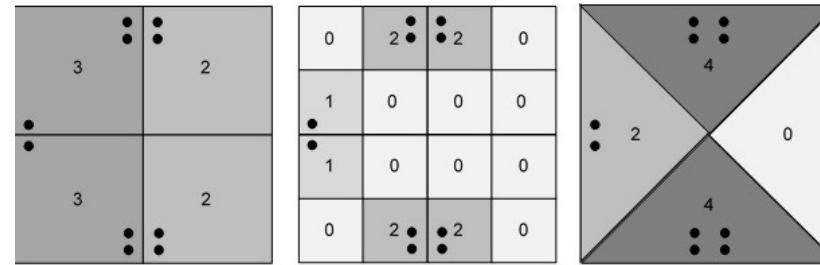


# RECAP



MAUP

## Modifiable Areal Unit Problem (MAUP)

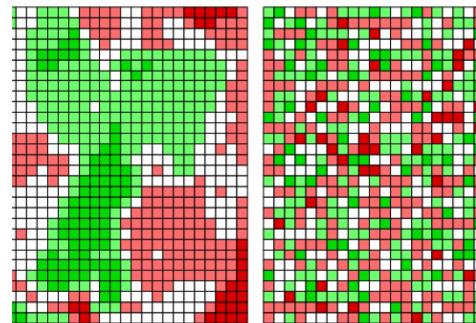


Source: <https://gistbok.ucgis.org/bok-topics/problems-scale-and-zoning>

The impact on aggregating data for a set of points to an original partitioning scheme (left panel) due to differences in scale (middle panel) and zoning (right panel). explanation.

SA

## Spatial Autocorrelation (SA)

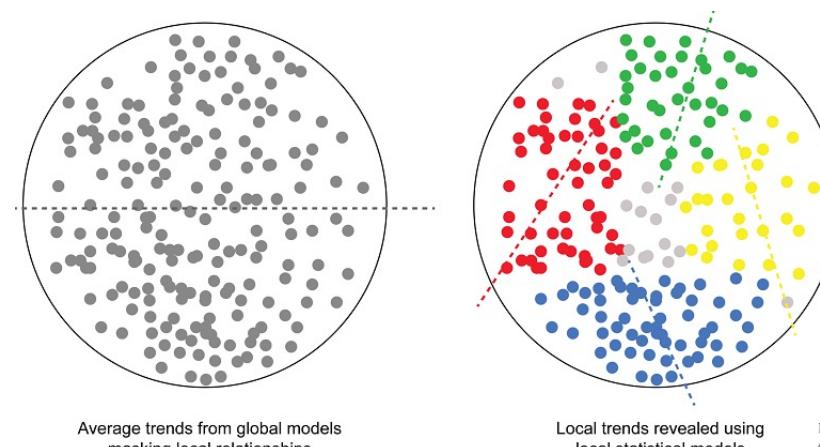


Source: <https://gistbok.ucgis.org/bok-topics/spatial-autocorrelation>

Everything is related to everything else, but nearby things are more related than distant things.



## Spatial Heterogeneity (SH)



Average trends from global models  
masking local relationships

Local trends revealed using  
local statistical models

N

Source: <https://gistbok.ucgis.org/bok-topics/geographically-weighted-regression-framework>

# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

**GWPCA:** *perform PCA analysis in a local way to reveal location-related principal components of variability*



MAUP

SA

SH



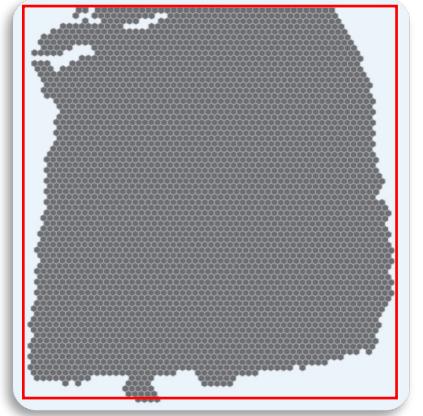
# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

**GWPCA:** perform PCA analysis in a local way to reveal location-related principal components of variability

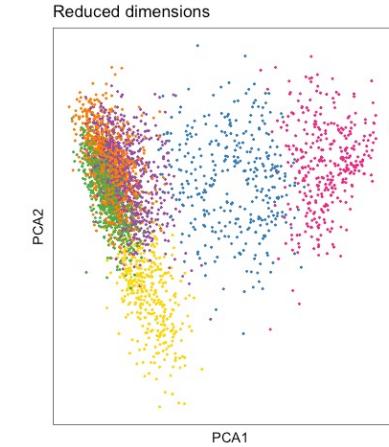
MAUP

SA

SH



GLOBAL  
-CLASSIC-  
PCA



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

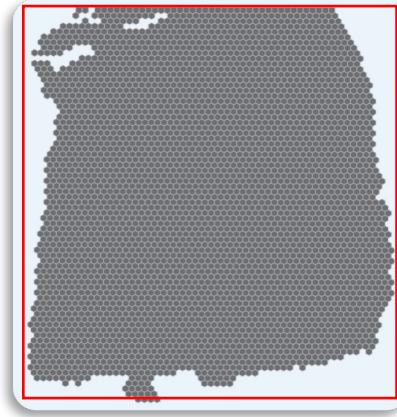
**GWPCA**: perform PCA analysis in a local way to reveal location-related principal components of variability



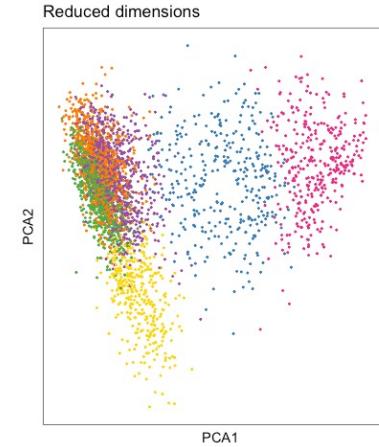
MAUP

SA

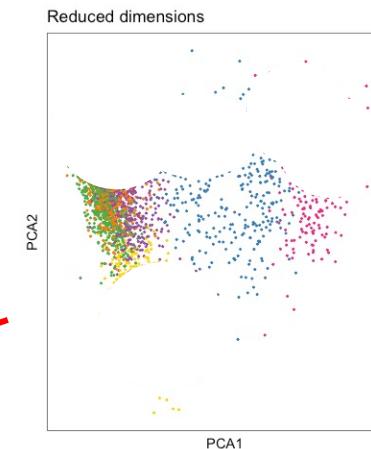
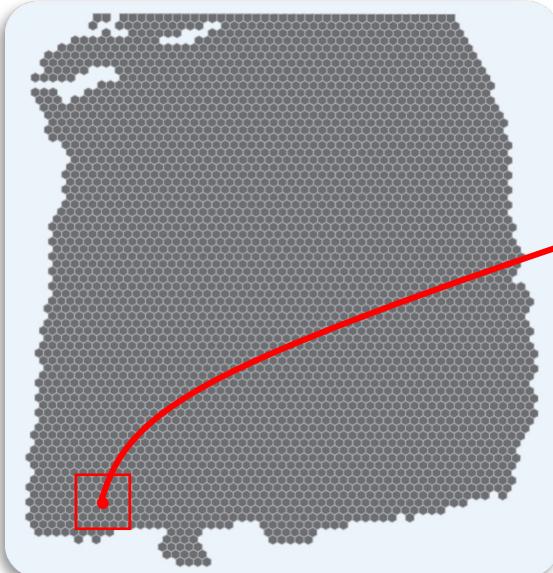
SH



GLOBAL  
-CLASSIC-  
PCA



GWPCA: “moving windows”



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

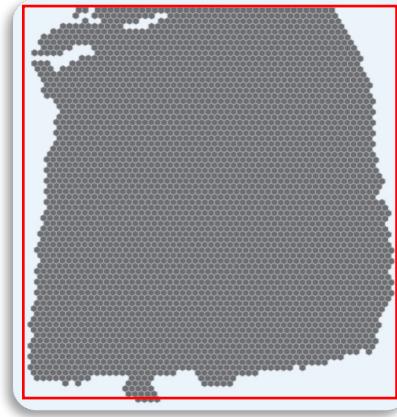
**GWPCA**: perform PCA analysis in a local way to reveal location-related principal components of variability



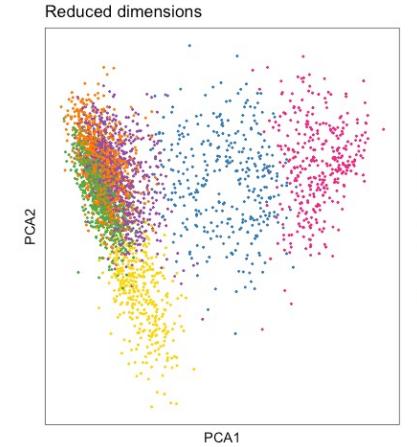
MAUP

SA

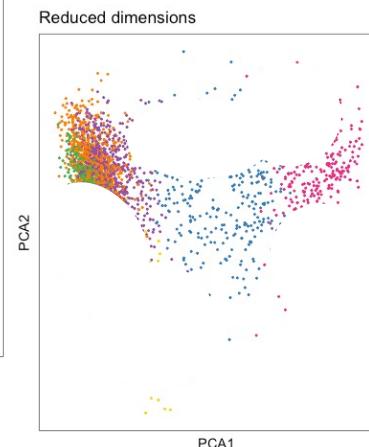
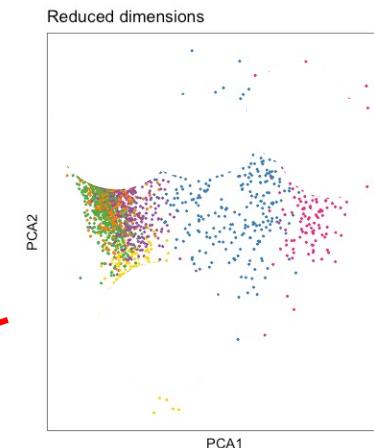
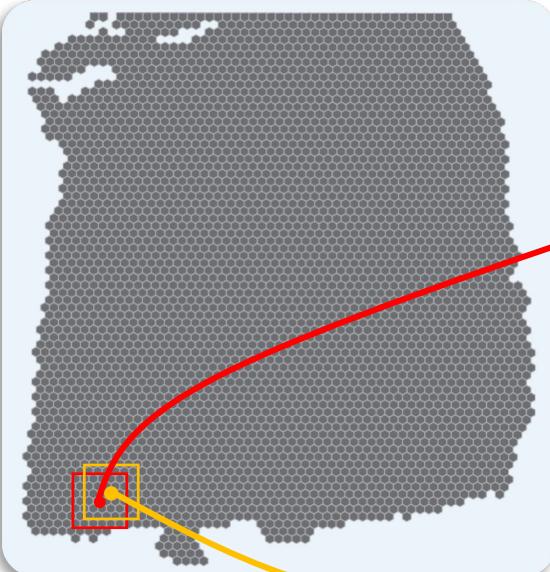
SH



GLOBAL  
-CLASSIC-  
PCA



GWPCA: “moving windows”



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

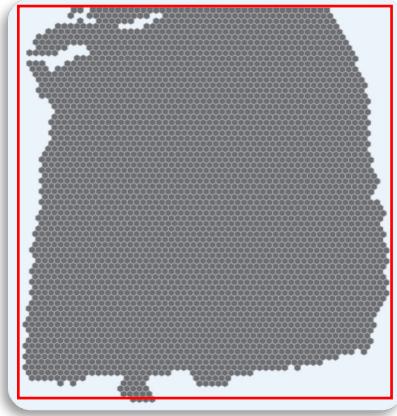
**GWPCA**: perform PCA analysis in a local way to reveal location-related principal components of variability



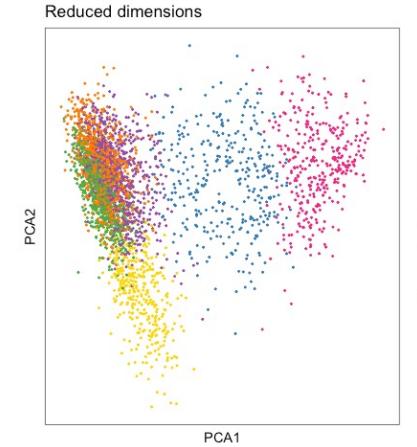
MAUP

SA

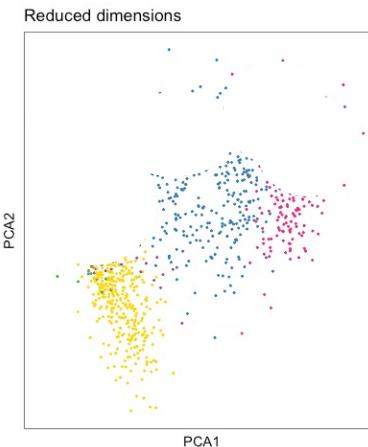
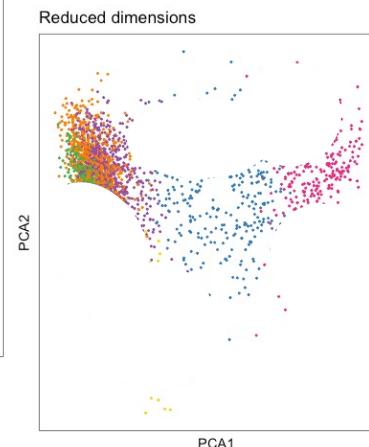
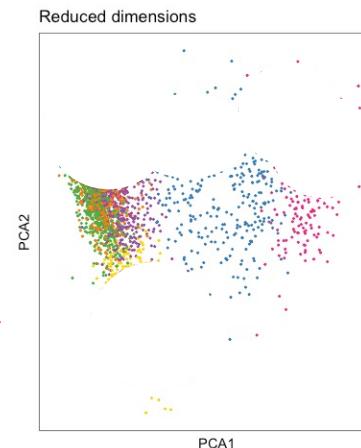
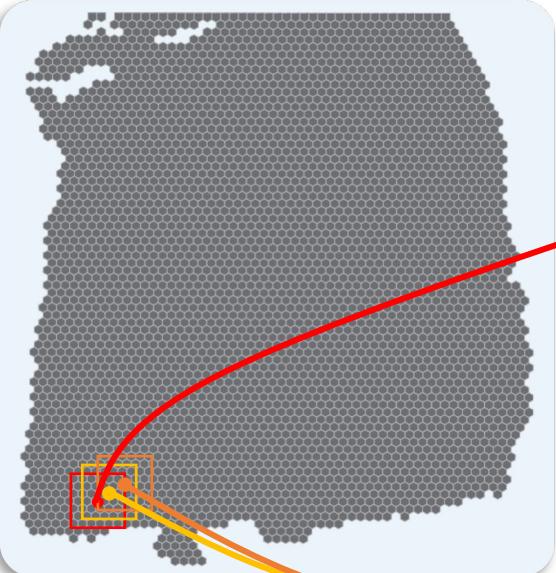
SH



GLOBAL  
-CLASSIC-  
PCA



GWPCA: “moving windows”



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

*GWPCA: perform PCA analysis in a local way to reveal location-related principal components of variability*

## WHICH SINGLE GENES DRIVE THE VARIABILITY IN EACH LOCATION?

MAUP

SA

SH



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

*GWPCA: perform PCA analysis in a local way to reveal location-related principal components of variability*



MAUP

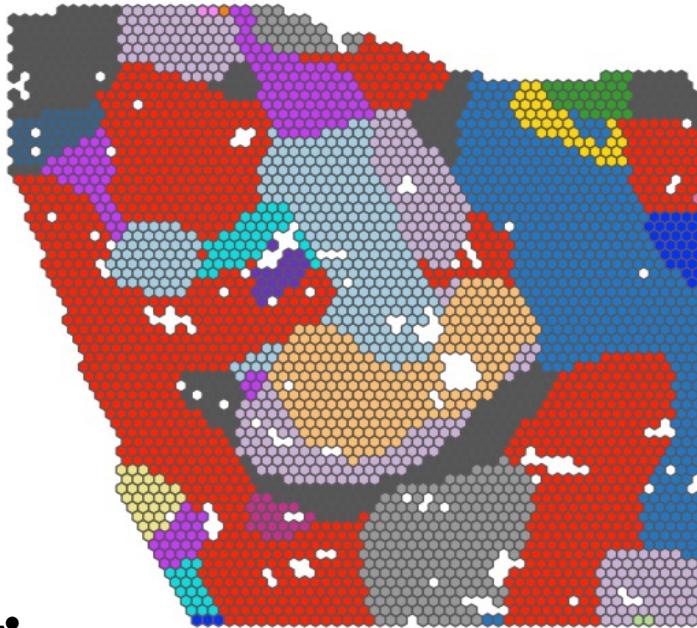
SA

SH



## WHICH SINGLE GENES DRIVE THE VARIABILITY IN EACH LOCATION?

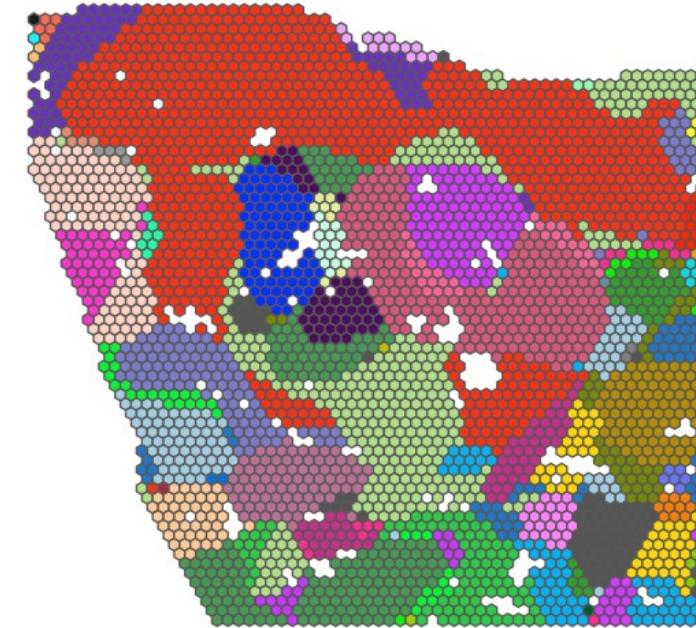
Leading Genes on PC1



Leading Genes

- ACPP
- AZGP1
- CFD
- CYP3A5
- FKBP5
- KLK2
- KLK3
- KLK4
- KRT5
- LTF
- MSMB
- MT1G
- NEFH
- NPY
- PSCA
- PTN
- RDH11
- SGK1
- SLC45A3
- VEGFA

Leading Genes on PC2



Leading Genes

- A2M
- ACPP
- APOD
- APOE
- BGN
- C7
- CCN1
- CD74
- CFB
- CFD
- CPE
- CXCL13
- DDIT4
- FBLN1
- FOS
- FXYD6
- SERPINF1
- SGK1
- SLC22A3
- SPON2
- SYNE1
- GLUL
- GPC3
- GRHL2
- HLA-DPA1
- HLA-DPB1
- HLA-DRA
- HSD11B1
- MT1G
- MT2A
- NCAPD3
- NEFH
- NELL2
- NPY
- PCP4
- PSCA
- PTGDS
- PTN
- KLK3
- KRT15
- KRT5
- LCN2
- LTF
- MSMB
- PTN
- TAGLN2
- TGM4
- TRPM8
- TSPAN1
- VWF
- WFDC2
- ZBTB16

# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

*GWPCA: perform PCA analysis in a local way to reveal location-related principal components of variability*

## WHICH SINGLE GENES DRIVE THE VARIABILITY IN EACH LOCATION?

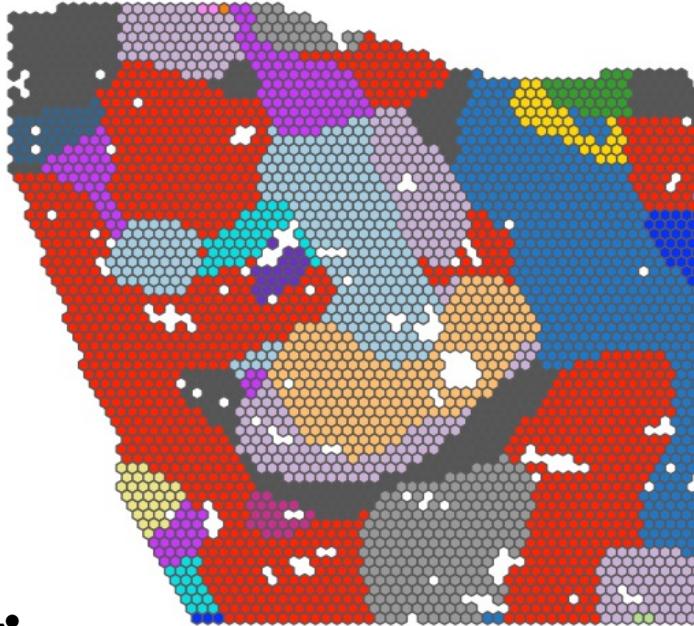
MAUP

SA

SH



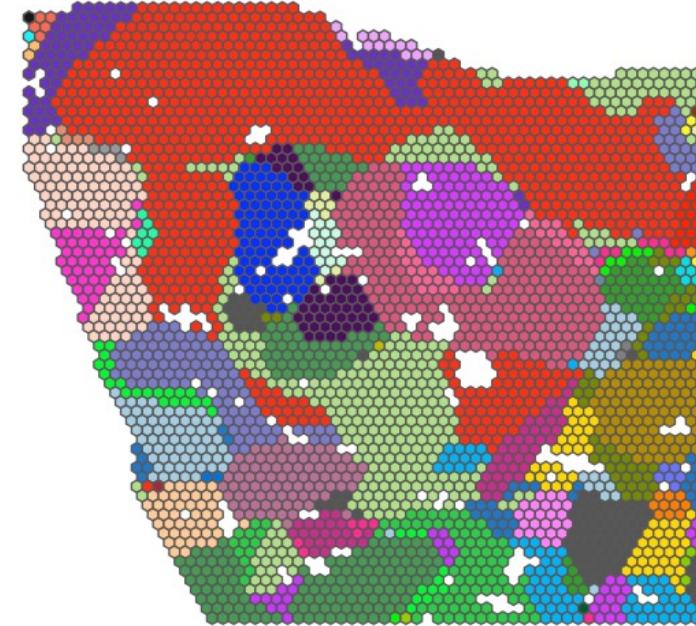
Leading Genes on PC1



Leading Genes

ACPP
AZGP1
CFD
CYP3A5
FKBP5
KLK2
KLK3
KLK4
KRT5
LTF
MSMB
MT1G
NEFH
NPY
PSCA
PTN
RDH11
SGK1
SLC45A3
VEGFA

Leading Genes on PC2



Leading Genes

A2M	GLUL	KLK3
ACPP	GPC3	KRT15
APOD	GRHL2	KRT5
APOE	HLA-DPA1	LCN2
BGN	HLA-DPB1	LTF
C7	HLA-DRA	MSMB
CCN1	HSD11B1	MT1G
CD74	MT2A	MT2A
CFB	HSP90AB1	NCAPD3
CFD	ID3	NEFH
CPE	IGF1	NELL2
CXCL13	IGFBP3	NPY
DDIT4	IGFBP5	PCP4
FBLN1	IGKC	PSCA
FOS	ITM2B	PTGDS
FXYD6	KLK2	PTN
SERPINF1	TAGLN2	
SGK1	TGM4	
SLC22A3	TRPM8	
SLC45A3	TSPAN1	
SPON2	VWF	
SYNE1	WFDC2	
	ZBTB16	

# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## WHAT PROCESSES/ PATHWAYS DO THEY AFFECT?

*Functional clustering using GWPCA PC1 with GSEA and MSigDB*

→ *GWPCA reveals the affected biology in the tissue areas*



MAUP

SA

SH



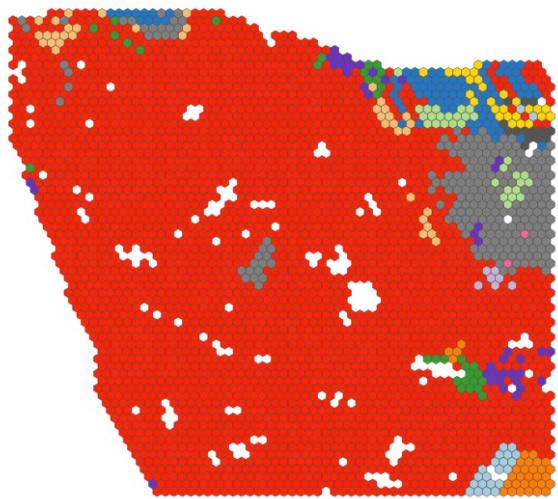
# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## WHAT PROCESSES/ PATHWAYS DO THEY AFFECT?

*Functional clustering using GWPCA PC1 with GSEA and MSigDB*

→ GWPCA reveals the affected biology in the tissue areas

MAUP



SA

SH

### HALLMARK GENES SET

- ALLOGRAFT\_REJECTION
- ANDROGEN\_RESPONSE
- ESTROGEN\_RESPONSE\_EARLY
- ESTROGEN\_RESPONSE\_LATE
- HYPOXIA
- IL6\_JAK\_STAT3\_SIGNALING
- INTERFERON\_ALPHA\_RESPONSE
- INTERFERON\_GAMMA\_RESPONSE
- KRAS\_SIGNALING\_DN
- OXIDATIVE\_PHOSPHORYLATION
- P53\_PATHWAY
- UNFOLDED\_PROTEIN\_RESPONSE
- NA



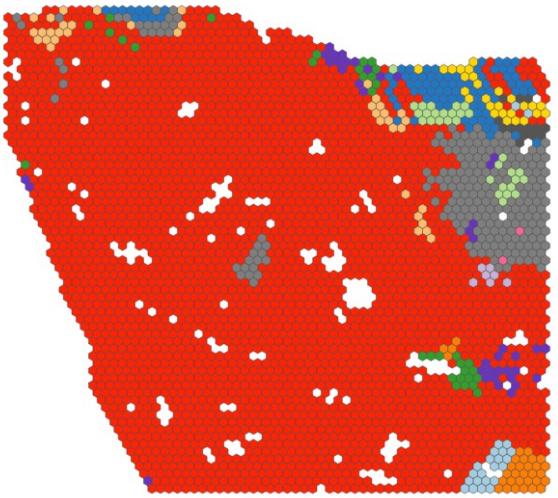
# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## WHAT PROCESSES/ PATHWAYS DO THEY AFFECT?

Functional clustering using GWPCA PC1 with GSEA and MSigDB

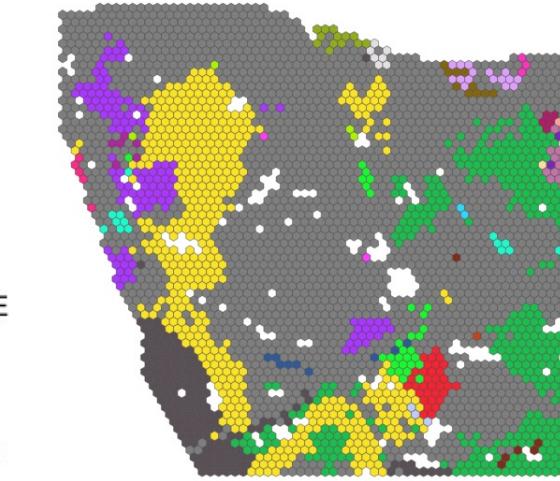
→ GWPCA reveals the affected biology in the tissue areas

MAUP



### HALLMARK GENES SET

- ALLOGRAFT\_REJECTION
- ANDROGEN\_RESPONSE
- ESTROGEN\_RESPONSE\_EARLY
- ESTROGEN\_RESPONSE\_LATE
- HYPOXIA
- IL6\_JAK\_STAT3\_SIGNALING
- INTERFERON\_ALPHA\_RESPONSE
- INTERFERON\_GAMMA\_RESPONSE
- KRAS\_SIGNALING\_DN
- OXIDATIVE\_PHOSPHORYLATION
- P53\_PATHWAY
- UNFOLDED\_PROTEIN\_RESPONSE
- NA



### GENE SETS RESPONDING TO CERTAIN ONCOGENIC PATHWAY BEHAVIOUR

■ AKT_UP_MTOR_DN --> UP	■ LTE2_UP --> UP
■ ALK_DN --> DN	■ MEK_UP --> UP
■ ATF2_S_UP --> DN	■ MTOR_UP --> DN
■ ATF2_S_UP --> UP	■ NOTCH_DN --> DN
■ ATF2_UP --> DN	■ P53_DN --> DN
■ BMI1_DN_MEL18_DN --> DN	■ P53_DN --> UP
■ CRX_DN --> DN	■ PTEN_DN --> DN
■ E2F1_UP --> DN	■ RB_DN --> DN
■ ERBB2_UP --> UP	■ RELA_DN --> DN
■ ESC_J1_UP_EARLY --> UP	■ RPS14_DN --> DN
■ ESC_V6.5_UP_EARLY --> DN	■ STK33_SKM_UP
■ IL15_UP --> DN	■ STK33_UP
■ IL2_UP --> UP	■ TBK1_DF_UP
■ IL21_UP --> DN	■ VEGF_A_UP --> DN
■ NA	■ NA

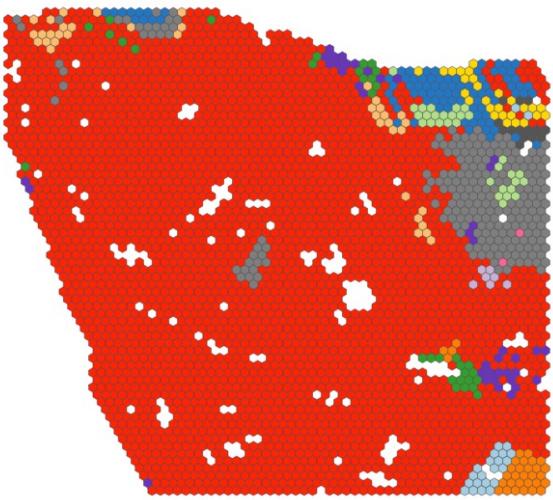
# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## WHAT PROCESSES/ PATHWAYS DO THEY AFFECT?

Functional clustering using GWPCA PC1 with GSEA and MSigDB

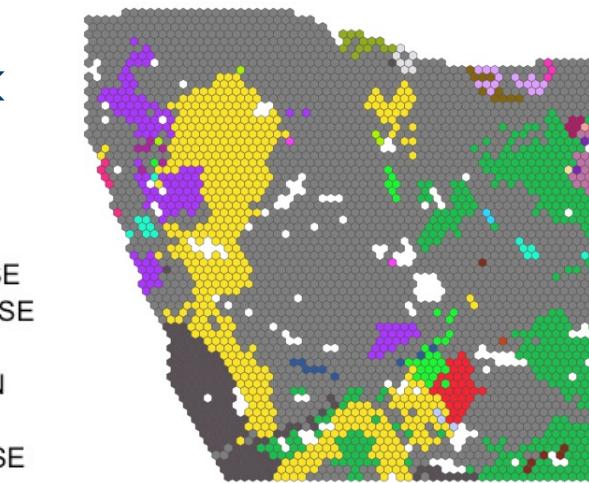
→ GWPCA reveals the affected biology in the tissue areas

MAUP



### HALLMARK GENES SET

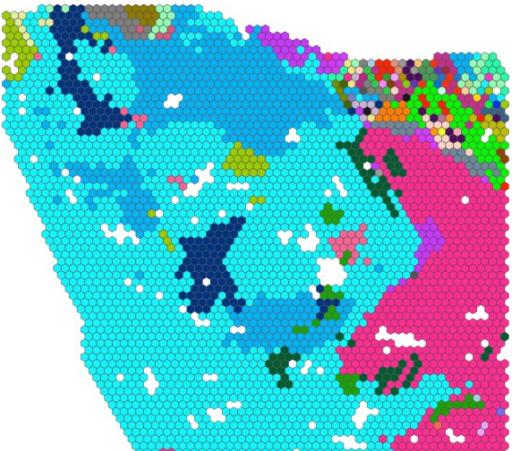
- ALLOGRAFT\_REJECTION
- ANDROGEN\_RESPONSE
- ESTROGEN\_RESPONSE\_EARLY
- ESTROGEN\_RESPONSE\_LATE
- HYPOXIA
- IL6\_JAK\_STAT3\_SIGNALING
- INTERFERON\_ALPHA\_RESPONSE
- INTERFERON\_GAMMA\_RESPONSE
- KRAS\_SIGNALING\_DN
- OXIDATIVE\_PHOSPHORYLATION
- P53\_PATHWAY
- UNFOLDED\_PROTEIN\_RESPONSE
- NA



### GENE SETS RESPONDING TO CERTAIN ONCOGENIC PATHWAY BEHAVIOUR

AKT_UP_MTOR_DN --> UP	LTE2_UP --> UP
ALK_DN --> DN	MEK_UP --> UP
ATF2_S_UP --> DN	MTOR_UP --> DN
ATF2_S_UP --> UP	NOTCH_DN --> DN
ATF2_UP --> DN	P53_DN --> DN
BMI1_DN_MEL18_DN --> DN	P53_DN --> UP
CRX_DN --> DN	PTEN_DN --> DN
E2F1_UP --> DN	RB_DN --> DN
ERBB2_UP --> UP	RELA_DN --> DN
ESC_J1_UP_EARLY --> UP	RPS14_DN --> DN
ESC_V6.5_UP_EARLY --> DN	STK33_SKM_UP
IL15_UP --> DN	STK33_UP
IL2_UP --> UP	TBK1.DF_UP
IL21_UP --> DN	VEGF_A_UP --> DN
NA	NA

### CURATED PATHWAYS



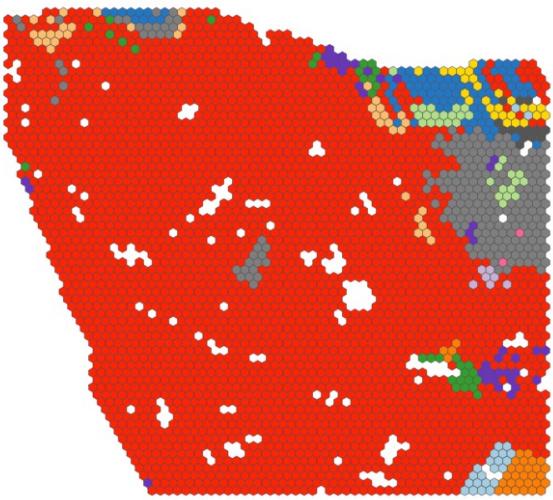
# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## WHAT PROCESSES/ PATHWAYS DO THEY AFFECT?

Functional clustering using GWPCA PC1 with GSEA and MSigDB

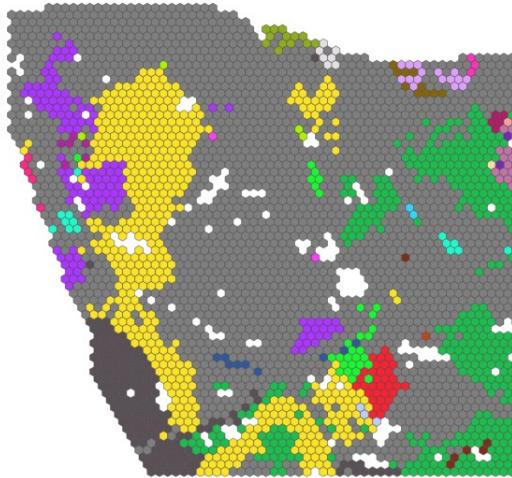
→ GWPCA reveals the affected biology in the tissue areas

MAUP



### HALLMARK GENES SET

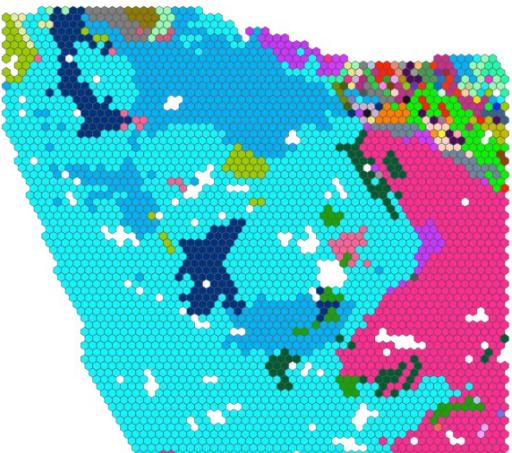
- ALLOGRAFT\_REJECTION
- ANDROGEN\_RESPONSE
- ESTROGEN\_RESPONSE\_EARLY
- ESTROGEN\_RESPONSE\_LATE
- HYPOXIA
- IL6\_JAK\_STAT3\_SIGNALING
- INTERFERON\_ALPHA\_RESPONSE
- INTERFERON\_GAMMA\_RESPONSE
- KRAS\_SIGNALING\_DN
- OXIDATIVE\_PHOSPHORYLATION
- P53\_PATHWAY
- UNFOLDED\_PROTEIN\_RESPONSE
- NA



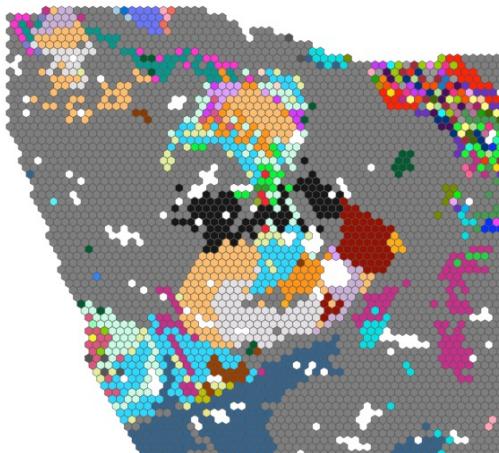
### GENE SETS RESPONDING TO CERTAIN ONCOGENIC PATHWAY BEHAVIOUR

AKT_UP_MTOR_DN --> UP	LTE2_UP --> UP
ALK_DN --> DN	MEK_UP --> UP
ATF2_S_UP --> DN	MTOR_UP --> DN
ATF2_S_UP --> UP	NOTCH_DN --> DN
ATF2_UP --> DN	P53_DN --> DN
BMI1_DN_MEL18_DN --> DN	P53_DN --> UP
CRX_DN --> DN	PTEN_DN --> DN
E2F1_UP --> DN	RB_DN --> DN
ERBB2_UP --> UP	RELA_DN --> DN
ESC_J1_UP_EARLY --> UP	RPS14_DN --> DN
ESC_V6.5_UP_EARLY --> DN	STK33_SKM_UP
IL15_UP --> DN	STK33_UP
IL2_UP --> UP	TBK1.DF_UP
IL21_UP --> DN	VEGF_A_UP --> DN
NA	NA

### CURATED PATHWAYS



### CANCER MODULES



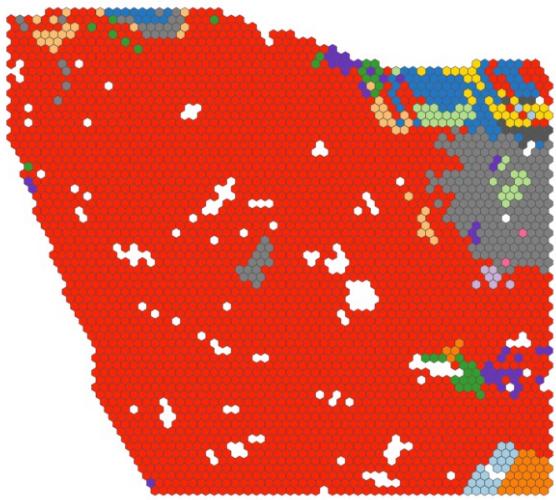
# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## WHAT PROCESSES/ PATHWAYS DO THEY AFFECT?

Functional clustering using GWPCA PC1 with GSEA and MSigDB

→ GWPCA reveals the affected biology in the tissue areas

MAUP

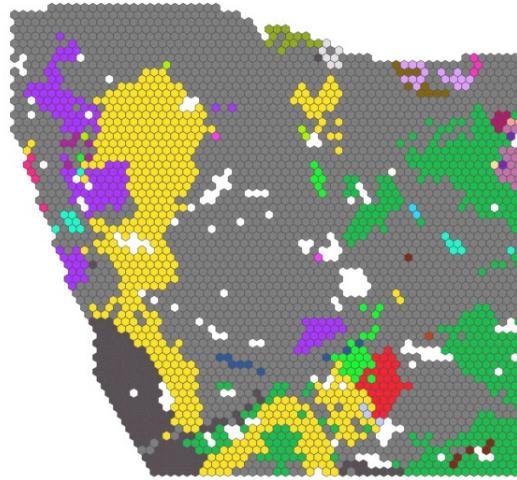


### HALLMARK GENES SET

- ALLOGRAFT\_REJECTION
- ANDROGEN\_RESPONSE
- ESTROGEN\_RESPONSE\_EARLY
- ESTROGEN\_RESPONSE\_LATE
- HYPOXIA
- IL6\_JAK\_STAT3\_SIGNALING
- INTERFERON\_ALPHA\_RESPONSE
- INTERFERON\_GAMMA\_RESPONSE
- KRAS\_SIGNALING\_DN
- OXIDATIVE\_PHOSPHORYLATION
- P53\_PATHWAY
- UNFOLDED\_PROTEIN\_RESPONSE
- NA



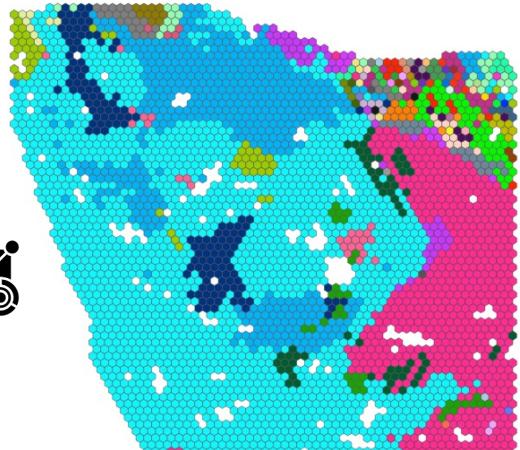
SA



### GENE SETS RESPONDING TO CERTAIN ONCOGENIC PATHWAY BEHAVIOUR

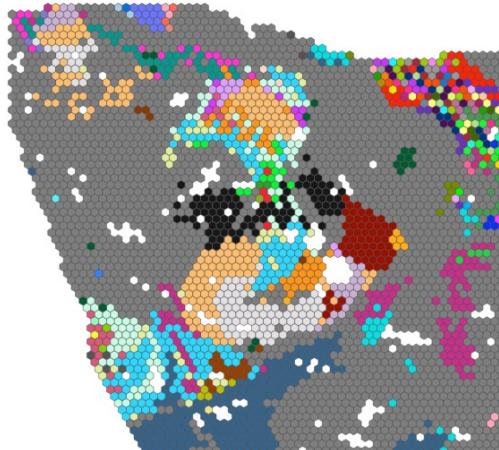
AKT_UP_MTOR_DN --> UP	LTE2_UP --> UP
ALK_DN --> DN	MEK_UP --> UP
ATF2_S_UP --> DN	MTOR_UP --> DN
ATF2_S_UP --> UP	NOTCH_DN --> DN
ATF2_UP --> DN	P53_DN --> DN
BMI1_DN_MEL18_DN --> DN	P53_DN --> UP
CRX_DN --> DN	PTEN_DN --> DN
E2F1_UP --> DN	RB_DN --> DN
ERBB2_UP --> UP	RELA_DN --> DN
ESC_J1_UP_EARLY --> UP	RPS14_DN --> DN
ESC_V6.5_UP_EARLY --> DN	STK33_SKM_UP
IL15_UP --> DN	STK33_UP
IL2_UP --> UP	TBK1.DF_UP
IL21_UP --> DN	VEGF_A_UP --> DN
NA	NA

SH

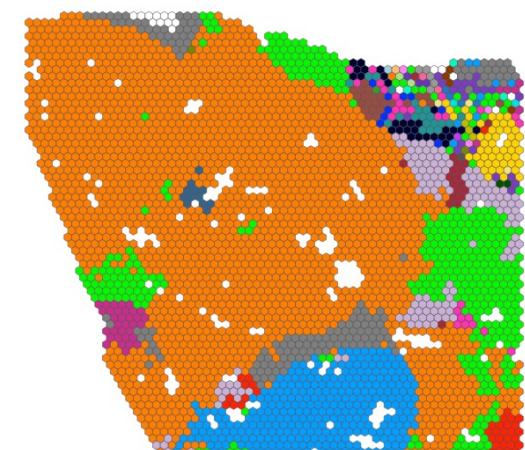


### CURATED PATHWAYS

### CANCER MODULES



### BIOLOGICAL PROCESSES



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## CAN WE IDENTIFY CANCER HOT SPOTS?

→ THE LOCAL DISCREPANCY SCORE SYSTEM

→ IDENTIFIES LOCATIONS THAT BEHAVE DIFFERENTLY TO THEIR SURROUNDINGS

MAUP

SA

SH



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

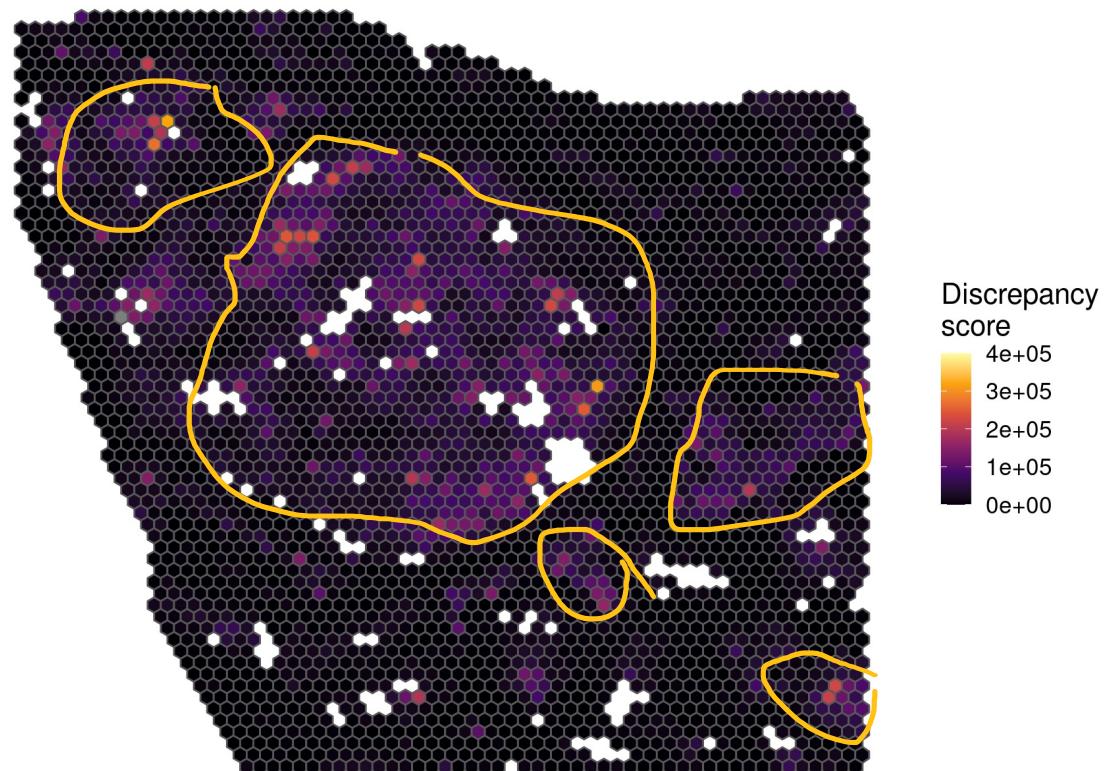
## CAN WE IDENTIFY CANCER HOT SPOTS?

→ THE LOCAL DISCREPANCY SCORE SYSTEM

→ IDENTIFIES LOCATIONS THAT BEHAVE DIFFERENTLY TO THEIR SURROUNDINGS

MAUP

Local PC Discrepancy



# GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS (GWPCA)

## CAN WE IDENTIFY CANCER HOT SPOTS?

→ THE LOCAL DISCREPANCY SCORE SYSTEM

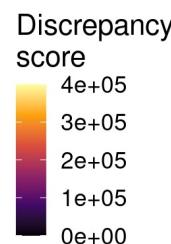
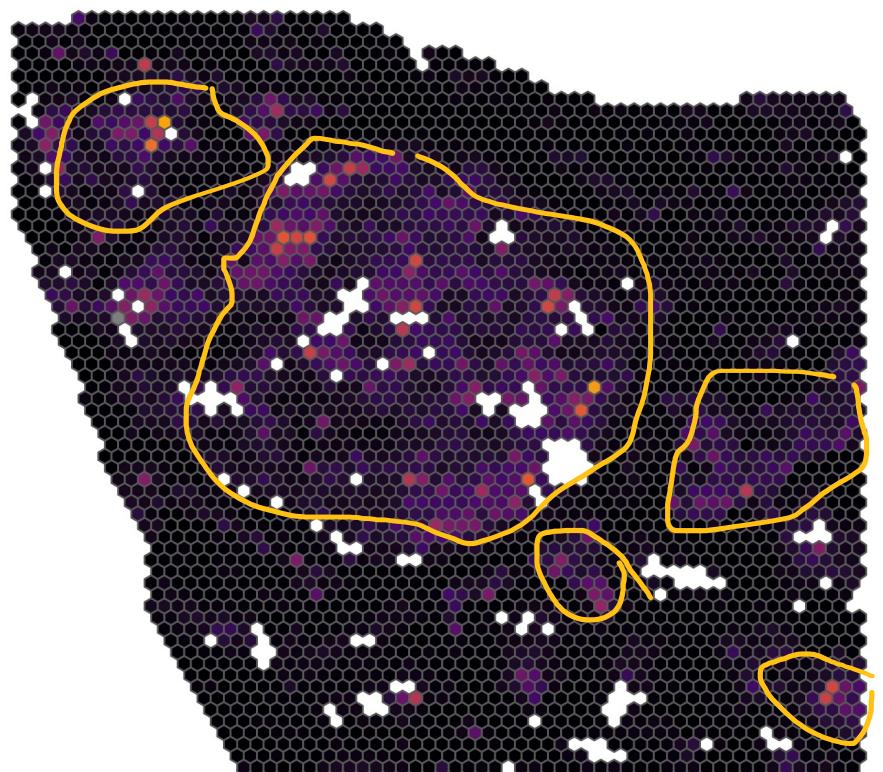
→ IDENTIFIES LOCATIONS THAT BEHAVE DIFFERENTLY TO THEIR SURROUNDINGS

MAUP

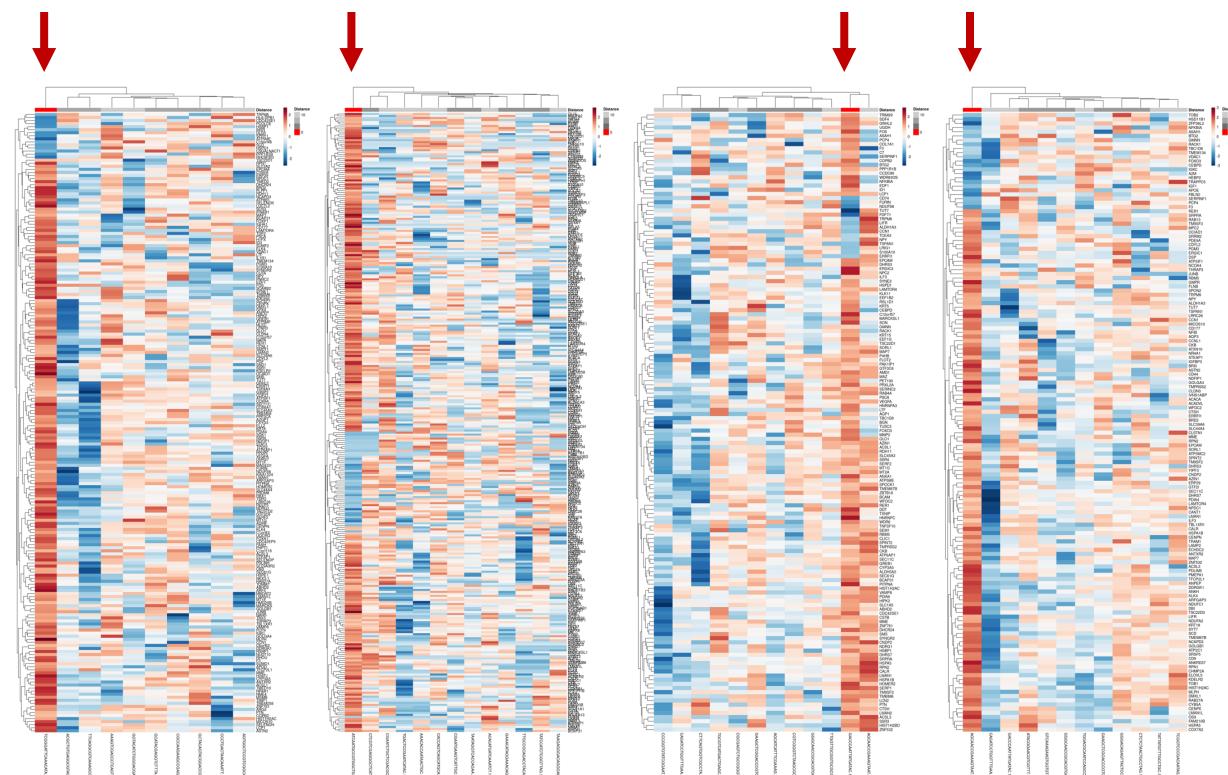
SA

SH

Local PC Discrepancy



Discrepancy VS neighbours HEATMAPS show the underlying gene expression differences



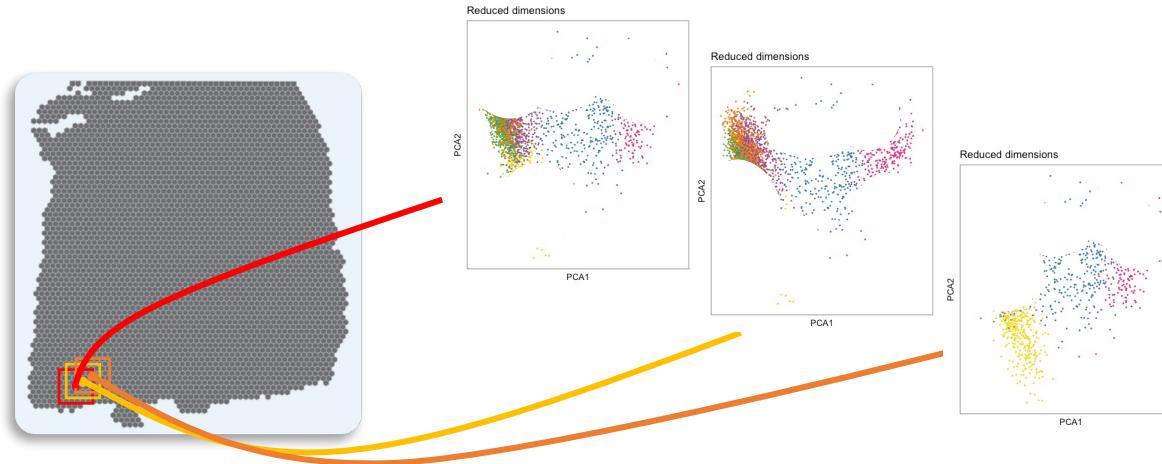


MAUP

SA

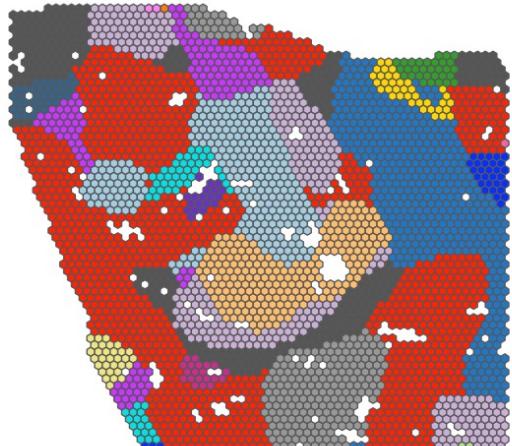
SH

## GWPCA: “moving windows”



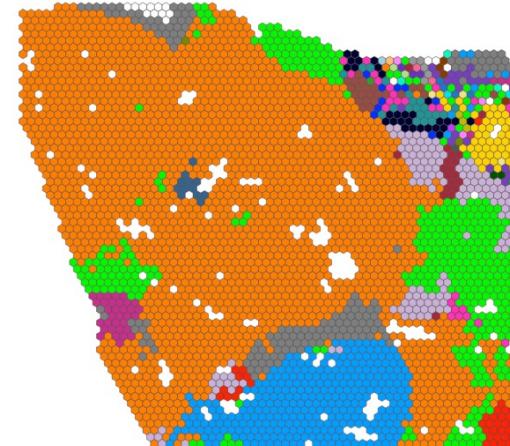
Which genes ...

Leading Genes on PC1



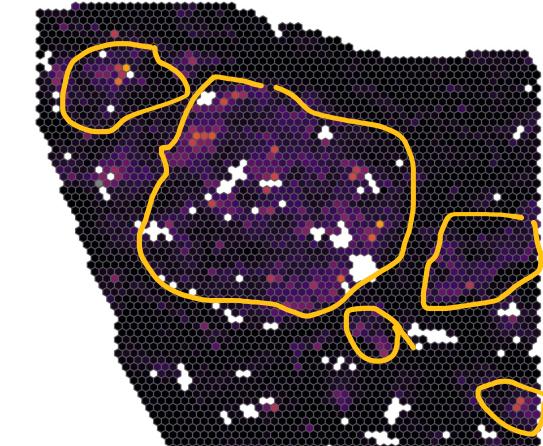
... do what and where

BIOLOGICAL PROCESSES



Which locations are “irregular”

Local PC Discrepancy



# AKNOWLEDGEMENTS

**Eleftherios Zormpas**



**Dr Simon J Cockell**



**Dr Rachel Queen**



**Prof. Alex Comber**



iSMB feedback form:



© ICBAM research group, Newcastle University, UK



Medical  
Research  
Council



MRC DiMeN  
Doctoral Training  
Partnership<sup>21</sup>