

Kernel-Based Testing for Single-Cell Differential Analysis

Anthony Ozier-Lafontaine[†], Bertrand Michel[†], Franck Picard^{*}

[†]Laboratoire de mathématiques Jean Leray, Nantes

^{*}Laboratoire Biologie et Modélisation de la Cellule, CNRS ENS-Lyon

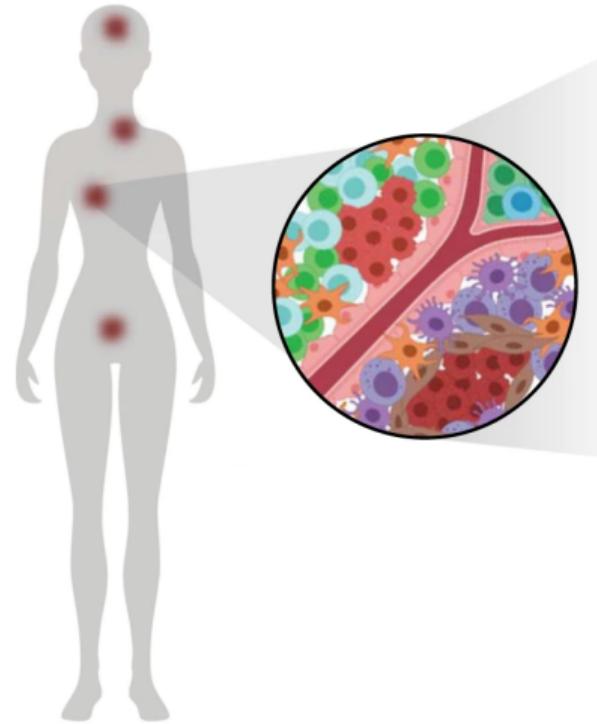
franck.picard@ens-lyon.fr

Outline

1. **The Single-Cell Revolution**
2. Introduction to kernel mean embedding
3. Introduction to kernel testing
4. ChemoResistance of Breast Cancer-Cells
5. Too many perspectives !

The cellular scale of biology

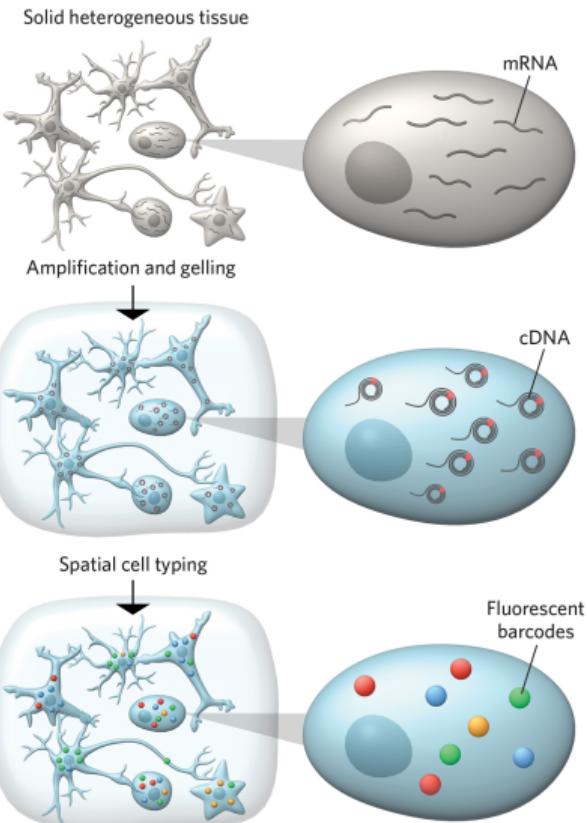
- Cells are the basic unit of structure and function in living organisms
- Cells are characterized by their 'types' that are diverse
- Physiology emerge thanks to complex interactions between different cell-types



[3]

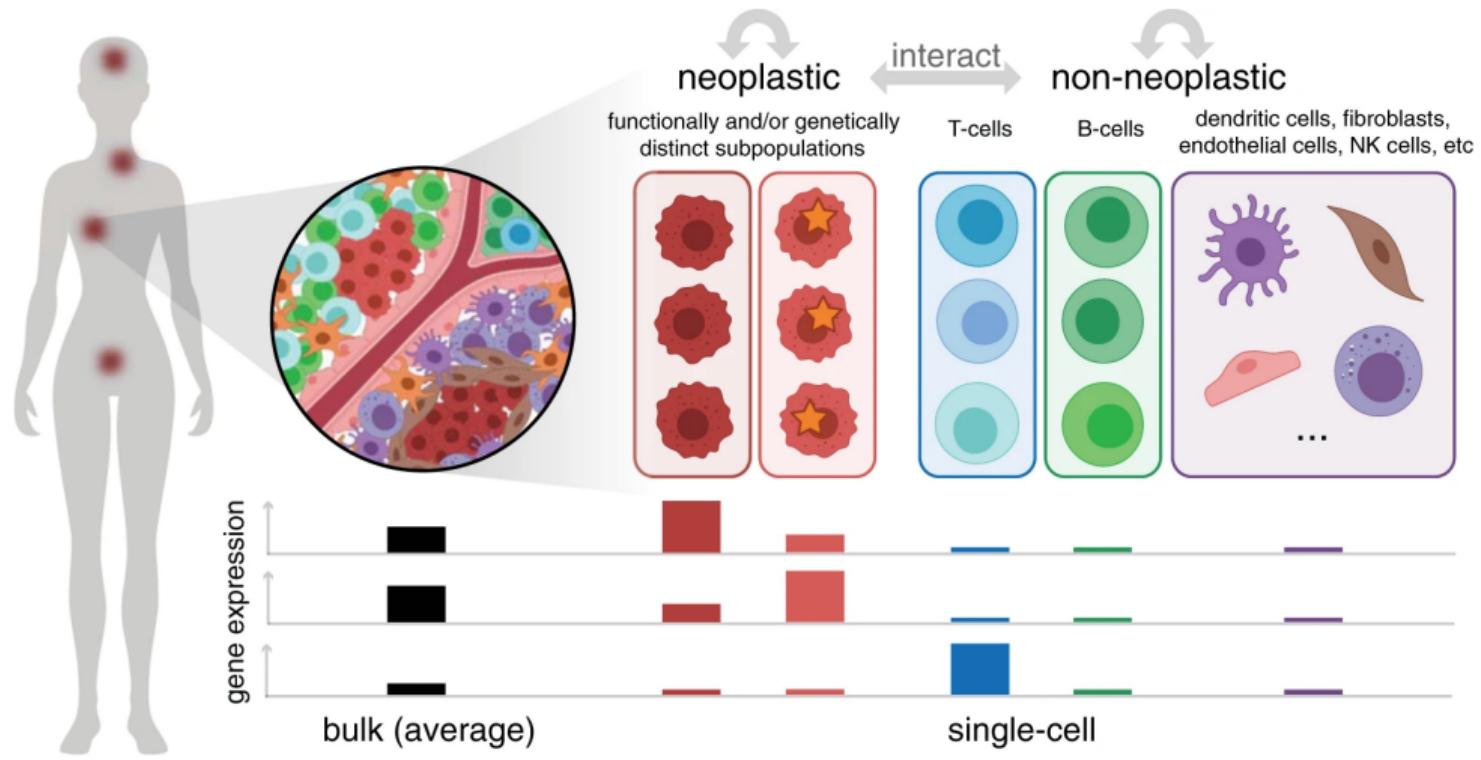
Cell functions are driven by molecular mechanisms

- Cells are characterized by the expression of their genes
- Gene expression is one key driver of cells phenotype
- The quantification of gene expression at the single-cell resolution is a technical challenge

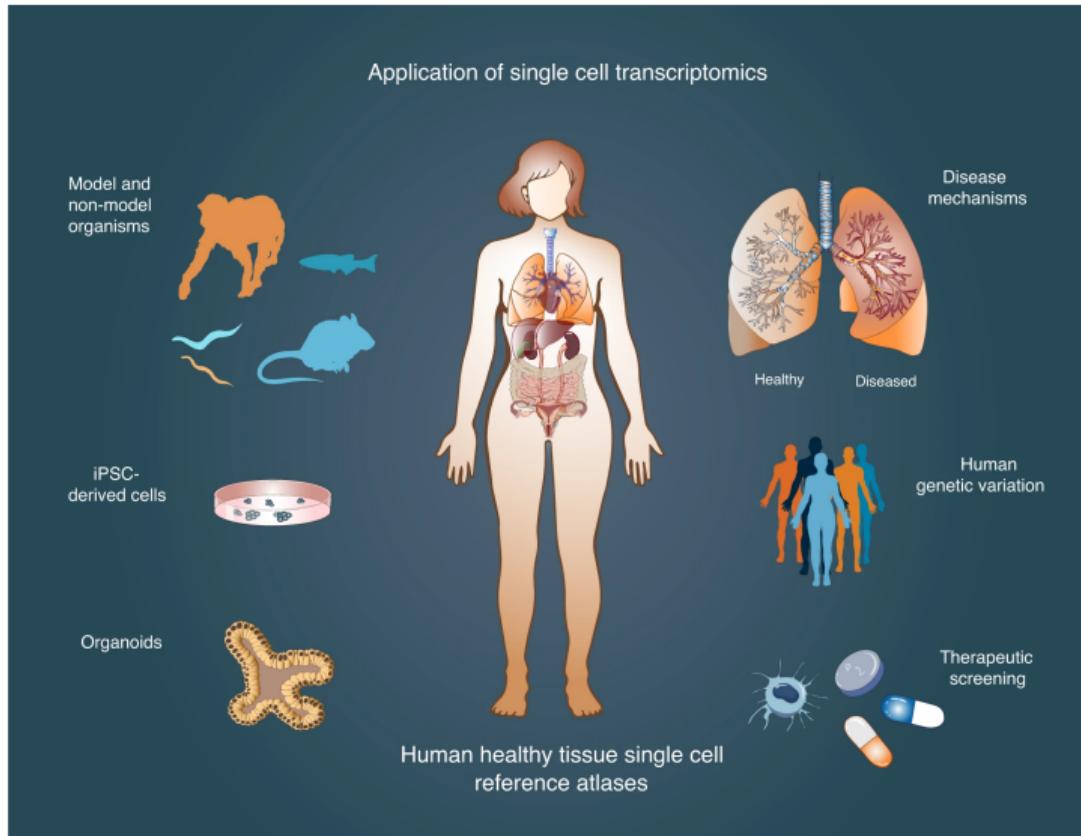


from *The Scientist*

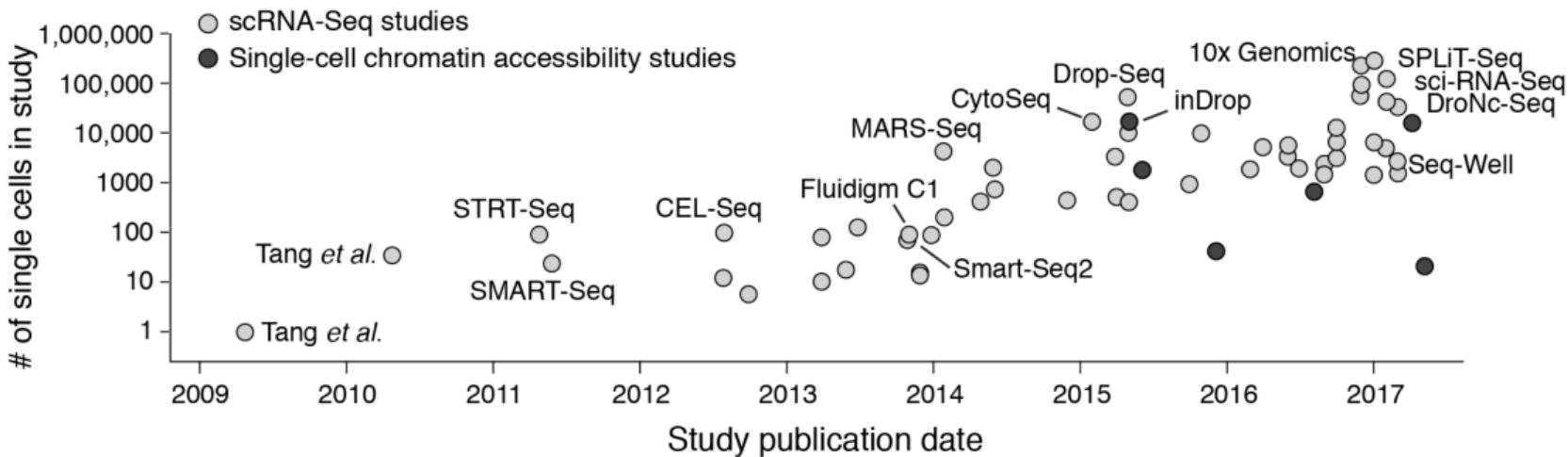
From bulk to distributions of gene expression



Potential of single-cell technologies



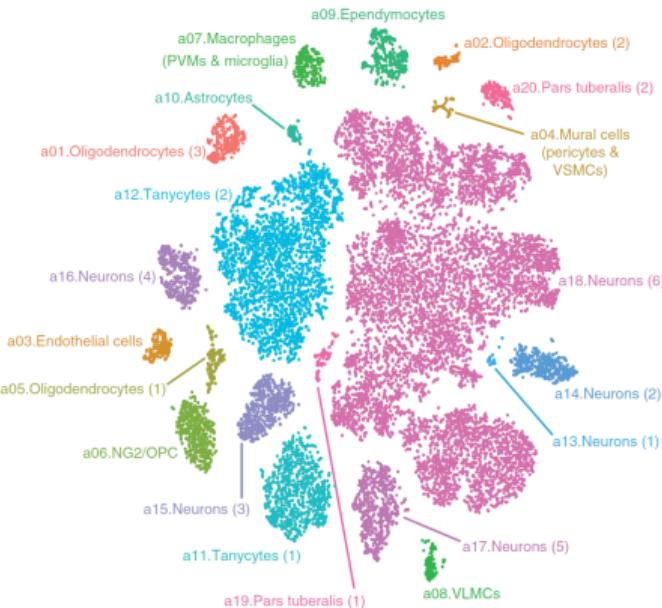
A timeline: produced data



[8]

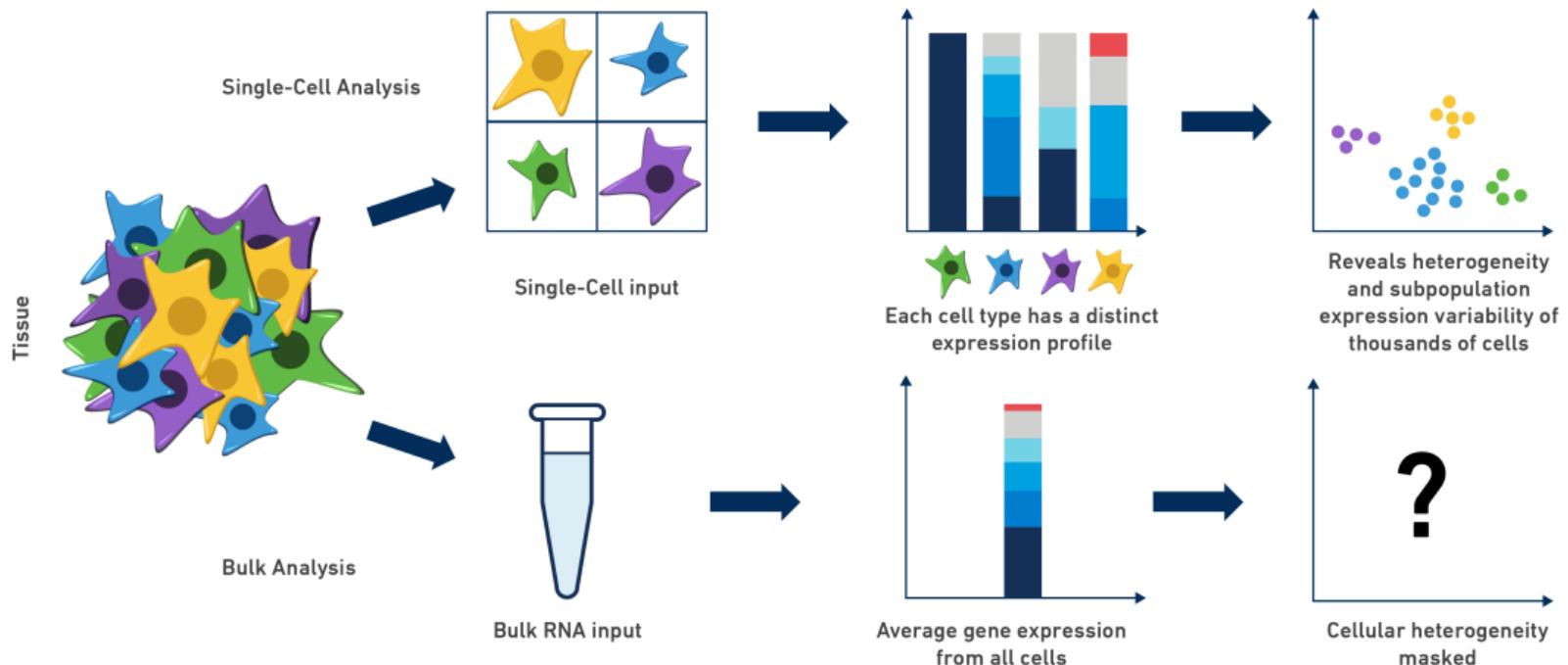
Machine Learning challenges

- Dimension Reduction / Visualization
- Clustering cell-type discovery (non supervised and semi supervised)
- Datasets alignments for non-matched samples
- Catch cells-ecosystems behaviors
- Simulation of fake data
- Data integration
- Statistical Testing (compare genes expression)



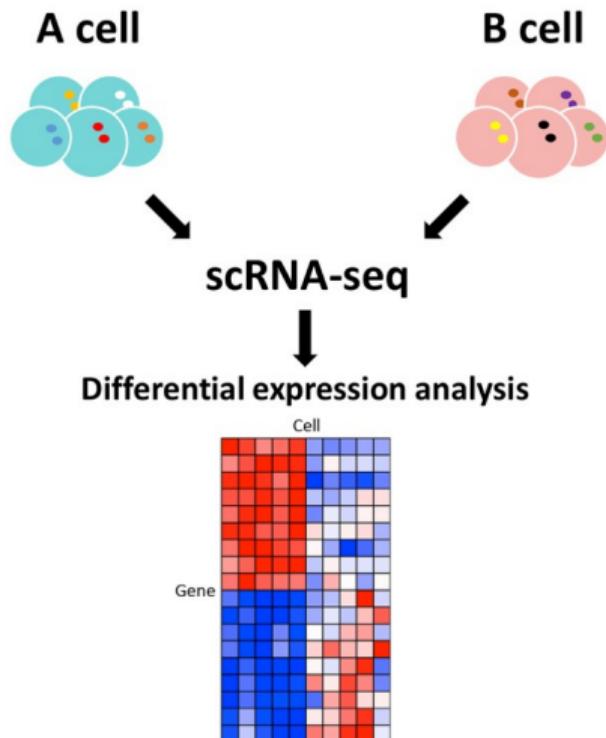
[2]

Single-Cell from a statistician's perspective



Differential Expression Analysis

- Context : compare biological conditions, identify genes (features)
- Task: two-sample test
- With single-cell data, many replicates ($n \sim 1 - 100,000$).
- Try non-parametrics !
- PhD of A. Ozier-Lafontaine with B. Michel (Nantes)

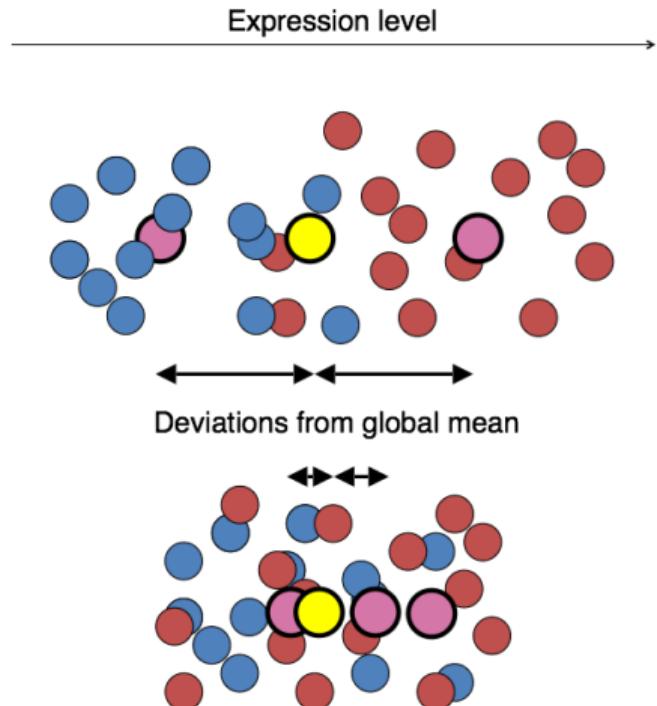


Statistical Setting: Linear Models

- Statistical setting: X_{ic}^g expression for gene g in condition i for cell c

$$\mathbb{E}(X_{ic}^g) = \mu_i^g$$

- Statistical testing of $\mathcal{H}_0^g : \{\mu_1^g = \mu_2^g\}$

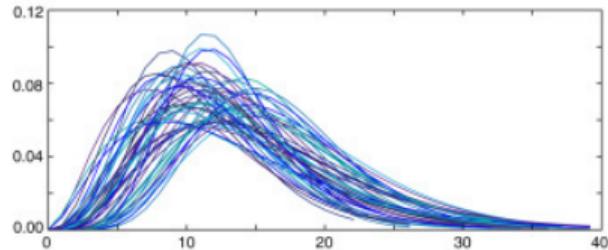
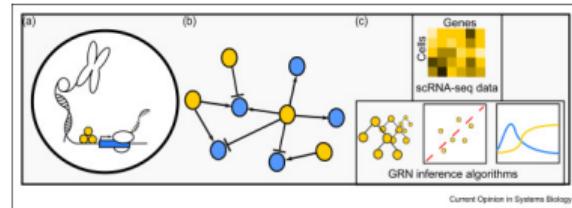


Strong dependencies and lots of data

- Gene Expressions are highly dependent
- Consider the multivariate model
- $\mathbf{X}_{ic} = [X_{ic}^1, \dots, X_{ic}^G], \mu_i = [\mu_i^1, \dots, \mu_i^G]$

$$\mathbb{E}(\mathbf{X}_{ic}) = \mu_i, \quad \mathbb{V}(\mathbf{X}_{ic}) = \Sigma$$

- Many cells $c \simeq 1 - 100,000$
- Σ can be inferred accurately
- Powerful Linear Multivariate Analysis



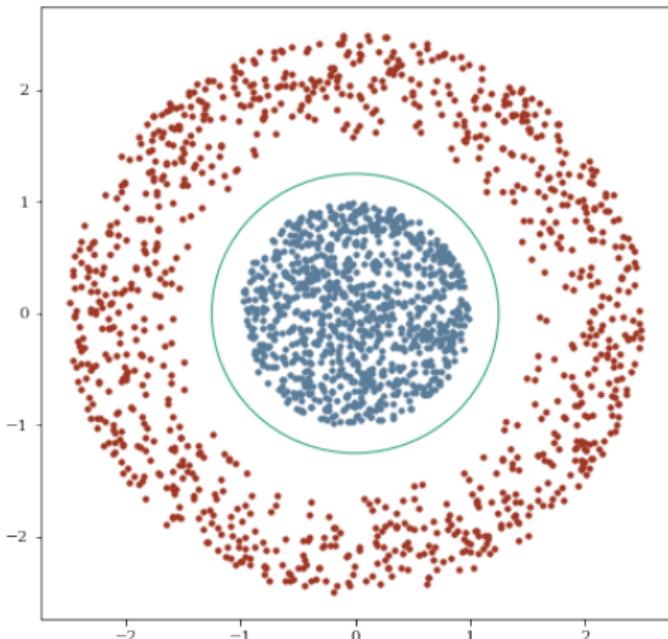
Distribution of gene expression across cells

Towards non-parametric testing

- Given the quantity of data, we can try non parametrics !
- $\mathbf{X}_1 \sim \mathbb{P}_1$ of size n_1 , $\mathbf{X}_2 \sim \mathbb{P}_2$ of size n_2
- Global distributional hypothesis

$$\mathcal{H}_0 : \left\{ \mathbb{P}_1 = \mathbb{P}_2 \right\}$$

- Can we find a non-linear separator to test the equality of distributions ?



Outline

1. The Single-Cell Revolution
2. **Introduction to kernel mean embedding**
3. Introduction to kernel testing
4. ChemoResistance of Breast Cancer-Cells
5. Too many perspectives !

Quick intro on kernel methods

- Kernel function : \mathcal{X} a measurable space:

$$k(\bullet, \bullet) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}.$$

- $k(\bullet, \bullet)$ is a positive definite kernel iff \mathbf{K} is symmetric and positive definite.

$$\forall (x_1, \dots, x_n) \in \mathcal{X}^n, \quad \mathbf{K} = [k(x_i, x_j)]_{i,j} \in \mathcal{M}_n(\mathbb{R})^n$$

$$\forall (c_1, \dots, c_n) \in \mathbb{R}^n, \sum_{i,j} c_i c_j k(x_i, x_j) \geq 0$$

Aronszajn Theorem

- Consider the so-called feature map function

$$\phi : \mathcal{X} \rightarrow \mathcal{H}_k$$

- $k(\bullet, \bullet)$ is positive definite iif there exists a Hilbert space \mathcal{H}_k from $\mathcal{X} \rightarrow \mathbb{R}$ and a feature map ϕ from $\mathcal{X} \rightarrow \mathcal{H}_k$, such that

$$\begin{aligned}\exists \phi &: \mathcal{X} \rightarrow \mathcal{H}_k \\ \phi_x(\bullet) &= \phi(x) = k(x, \bullet) \\ \forall (x, x') \in \mathcal{X}^2 &: k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}_k}.\end{aligned}$$

- Interpretation: given a p.d.k. $k(\bullet, \bullet)$, one can find ϕ and \mathcal{H}_k (unique) so that the relation is verified.

Aronszajn Theorem and the reproducing property

- A positive definite kernel k defines a space of functions \mathcal{H}_k that is a Reproducing Kernel Hilbert Space (RKHS)
- A RKHS has 2 important properties

$$i) \quad \forall x \in \mathcal{X}, \quad k(x, \bullet) : x' \mapsto k(x, x') \in \mathcal{H}_k$$

$$ii) \quad \forall f \in \mathcal{H}_k, x \in \mathcal{X}, \quad f(x) = \langle f, k(x, \bullet) \rangle_{\mathcal{H}_k}$$

- The kernel k represents point evaluation
- If $f(x') = k(x', \bullet)$ for some $x' \in \mathcal{H}_k$, then we obtain the reproducing property:

$$\langle k(x, \bullet), k(x', \bullet) \rangle_{\mathcal{H}_k} = k(x, x')$$

- Intuitively, functions in RKHS are rather smooth

Consequence of the Reproducing property

- The reproducing property comes directly from the Riesz representation theorem: if $\mathcal{A} : \mathcal{H} \rightarrow \mathbb{R}$ is a bounded linear operator in a HS \mathcal{H} , there exists $g_{\mathcal{A}} \in \mathcal{H}$ such that

$$\mathcal{A}[f] = \langle f, g_{\mathcal{A}} \rangle_{\mathcal{H}}, \quad f \in \mathcal{H}$$

- By definition of the RKHS, $f(x)$ is a linear operator in \mathcal{H} : kernel methods are linear methods (in \mathcal{H})
- The Riesz representation theorem ensures that for any $x \in \mathcal{X}$, we can find a representer of the evaluation $f(x)$
- When considering a kernel k we are sure that the point $x \in \mathcal{X}$ is implicitly represented by $\phi(x) = k(x, \bullet)$ in \mathcal{H} .
- Can we define a representer for distributions ?

Interpretations of Aronszajn theorem

- $k(\bullet, \bullet)$ is a shortcut to avoid the explicit determination of ϕ
- Kernelized statistical methods replace inner products by a kernel evaluation.
- It suffices to have a p.d.k that measures pair-wise similarities to run a statistical method on any kind of data
- We will consider the Gaussian kernel (many other possibilities)

$$k_\sigma(x, x') = \exp\left(-\frac{1}{2\sigma^2}\|x - x'\|_2^2\right)$$

Embedding distributions

- If we consider $\mathcal{M}_+^1(\mathcal{X})$ the space of probability measures over a measurable space \mathcal{X} , and \mathbb{P} a probability measure
- Define the representer $\mu_{\mathbb{P}}$ of \mathbb{P} in \mathcal{H}_k , such that

$$\begin{aligned}\mu_{\mathbb{P}} : \mathcal{M}_+^1(\mathcal{X}) &\rightarrow \mathcal{H}_k \\ \mathbb{P} &\rightarrow \mu_{\mathbb{P}} = \int k(x, \bullet) d\mathbb{P}(x)\end{aligned}$$

- $\mu_{\mathbb{P}}$ is called the mean embedding of distribution \mathbb{P} :

$$\mu_{\mathbb{P}} = \mathbb{E}_{X \sim \mathbb{P}} (\phi(X))$$

Embedding distributions

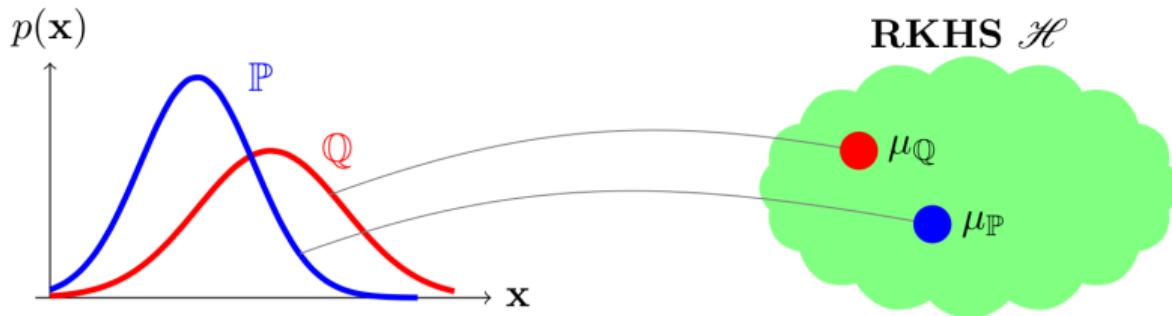


Figure 3.2: Embedding of marginal distributions: Each distribution is mapped into a reproducing kernel Hilbert space (RKHS) via an expectation operation.

from Muandet et al. 2017 [7]

Properties of the mean embedding

- Condition under which the mean embedding exists and belongs to \mathcal{H} :

$$\text{if } \mathbb{E}_{X \sim \mathbb{P}}(k(X, X)^{1/2}) < \infty$$

$$\text{then } \mu_{\mathbb{P}} \in \mathcal{H}$$

$$\mathbb{E}_{X \sim \mathbb{P}}(f(X)) = \langle f, \mu_{\mathbb{P}} \rangle_{\mathcal{H}}$$

- Reproducing property of the expectation operation in a RKHS
- It allows us to compute the expectation of a function f in the RKHS wrt \mathbb{P} by mean of an inner product between f and $\mu_{\mathbb{P}}$

Kernel Covariance Operators

- Represent distribution beyond the mean embedding
- Denote by $\Sigma_{\mathbb{P}}$ a Hilbert Schmidt operator

$$\text{if } \mathbb{E}_{X \sim \mathbb{P}}(k(X, X)^{1/2}) < \infty$$

- Then there exists an unique Hilbert-Schmidt operator $\Sigma_{\mathbb{P}}$ such that for $g, h \in \mathcal{H}_k$, we have:

$$\text{Cov}(g(y), h(y)) = \langle g, \Sigma_{\mathbb{P}} h \rangle_{\mathcal{H}_k}$$

- The kernel operator provides information on the covariance of the feature maps:

$$\Sigma_{\mathbb{P}} = \mathbb{E}_{X \sim \mathbb{P}} [(\phi(X) - \mu_{\mathbb{P}}) \otimes (\phi(X) - \mu_{\mathbb{P}})]$$

Outline

1. The Single-Cell Revolution
2. Introduction to kernel mean embedding
3. **Introduction to kernel testing**
4. ChemoResistance of Breast Cancer-Cells
5. Too many perspectives !

Metric between distributions

- Testing H_0 requires a metric between distributions

$$\mathcal{H}_0 : \left\{ \mathbb{P}_1 = \mathbb{P}_2 \right\}$$

- Let \mathcal{F} be a space of functions from \mathcal{X} to \mathbb{R}
- Define the metric

$$\mathcal{M}_{\mathcal{F}}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{f \in \mathcal{F}} |\mathbb{E}_{\mathbb{P}_1} f(x) - \mathbb{E}_{\mathbb{P}_2} f(x)|,$$

- $\mathcal{M}_{\mathcal{F}}(\mathbb{P}_1, \mathbb{P}_2)$ fulfills the separability property:

$$\mathbb{P}_1 = \mathbb{P}_2 \quad \Leftrightarrow \quad \mathcal{M}_{\mathcal{F}}(\mathbb{P}_1, \mathbb{P}_2) = 0.$$

Towards the Maximum Mean Discrepancy

- If \mathcal{F} is too large the metric is difficult to compute
- $\mathcal{F} = \{\mathbb{1}_{(-\infty, t]} \mid t \in \mathbb{R}\}$ gives the Kolmogorov distance
- $\mathcal{F} = \{f \mid \sup_{x \in \mathcal{X}} |f(x)| \leq 1\}$ gives the total variation distance.
- If $\mathcal{F} \subset \mathcal{H}$ and $\|f\|_{\mathcal{H}} = 1$, we obtain the maximal mean discrepancy

$$\mathcal{M}_{\mathcal{F}}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{\substack{f \in \mathcal{H} \\ \|f\|_{\mathcal{H}}=1}} |\mathbb{E}_{X \sim \mathbb{P}_1} f(X) - \mathbb{E}_{X \sim \mathbb{P}_2} f(X)|$$

$$= \sup_{\substack{f \in \mathcal{H} \\ \|f\|_{\mathcal{H}}=1}} |\langle f, \mu_1 - \mu_2 \rangle_{\mathcal{H}_k}|$$

$$\text{MMD}(\mathbb{P}_1, \mathbb{P}_2) = \|\mu_1 - \mu_2\|_{\mathcal{H}_k}$$

Computing the empirical MMD

- Embed the observations in \mathcal{H}_k and define the empirical mean embeddings

$$\hat{\mu}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \phi(X_{1,i}) \quad \hat{\mu}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \phi(X_{2,i})$$

- Compute the empirical MMD as a test statistic

$$\begin{aligned}\widehat{\text{MMD}}^2 &= \|\hat{\mu}_2 - \hat{\mu}_1\|_{\mathcal{H}}^2 \\ &= \frac{1}{n_1(n_1-1)} \sum_{i \neq j} k(X_{1,i}, X_{1,j}) + \frac{1}{n_2(n_2-1)} \sum_{i \neq j} k(X_{2,i}, X_{2,j}) \\ &\quad - \frac{2}{n_1 n_2} \sum_{i,j} k(X_{1,i}, X_{2,j})\end{aligned}$$

Connection with Multivariate Testing

- Consider the Multivariate Gaussian Model

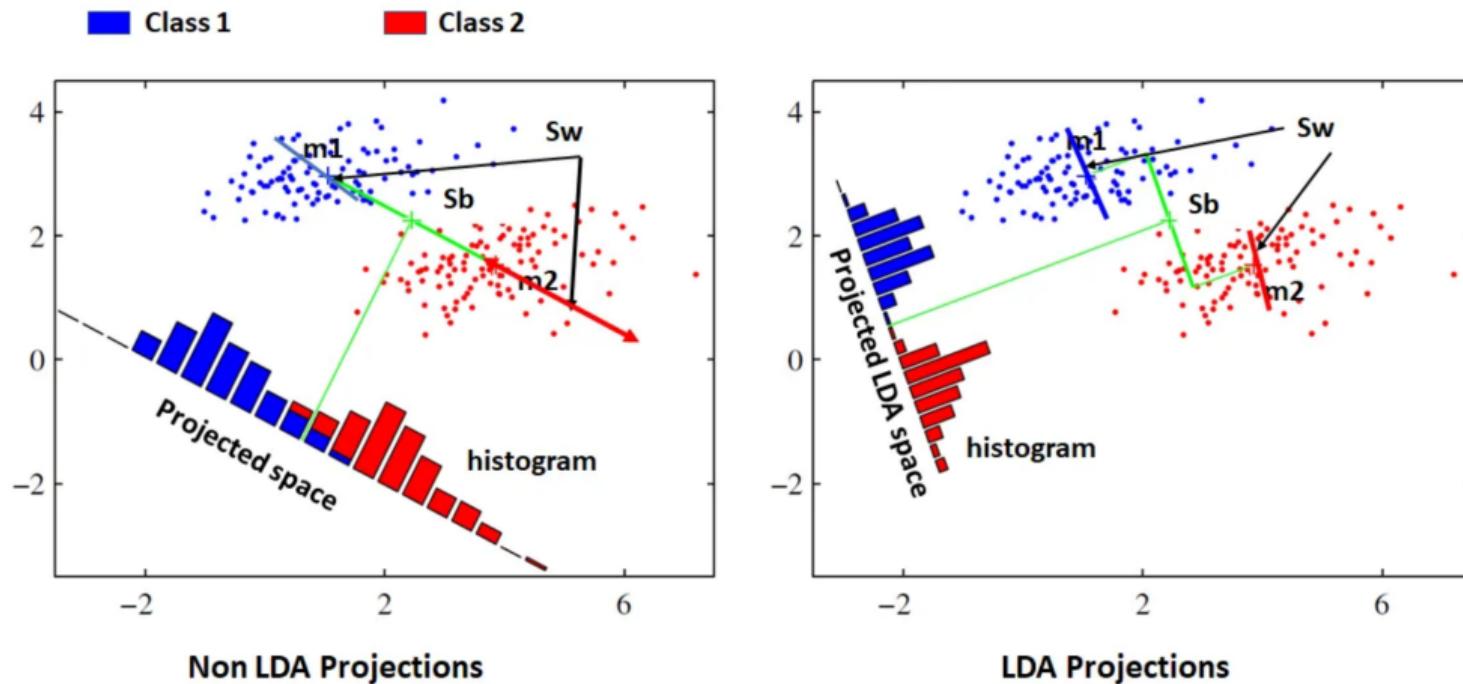
$$\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1), \quad \mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2), \quad (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathbb{R}^p$$

- Test $\mathcal{H}_0 : \{\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2\}$ using the Hotelling T^2 statistic

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} \|\Sigma_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\|_2^2, \quad \Sigma_W = \frac{n_1}{n} \Sigma_1 + \frac{n_2}{n} \Sigma_2$$

- Need to regularize Σ_W (Ridge, truncation)

Connection with Linear Discriminant Analysis



LDA vs Non LDA Projections from [TDS](#)

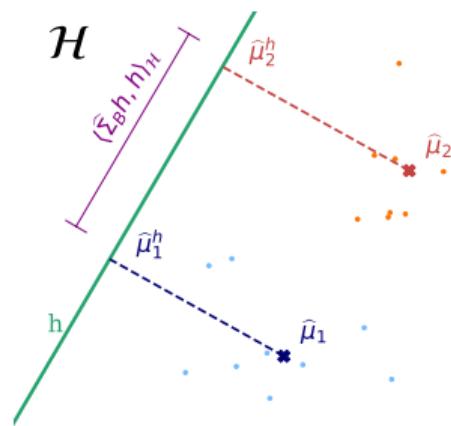
Geometric intuition for MMD

- The MMD is linked to the between-group covariance

$$\widehat{\Sigma}_B = \frac{n_1 n_2}{n^2} \left(\widehat{\mu}_2 - \widehat{\mu}_1 \right)^{\otimes 2}$$

- If h is an axis in \mathcal{H} , the distance between the projections over h is:

$$\left\langle h, \widehat{\Sigma}_B h \right\rangle_{\mathcal{H}} = \frac{n_1 n_2}{n} \left\langle h, \widehat{\mu}_1 - \widehat{\mu}_2 \right\rangle_{\mathcal{H}}^2$$



Limitations of the MMD

- Define the within-group covariances $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$

$$\widehat{\Sigma}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \left(\phi(X_{1,i}) - \widehat{\mu}_1 \right)^{\otimes 2}, \quad \widehat{\Sigma}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \left(\phi(X_{2,i}) - \widehat{\mu}_2 \right)^{\otimes 2}$$

- How to account for within-group covariances $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$?
- How to normalize the MMD according to the residual variability ?

Geometric intuition for Kernel FDA

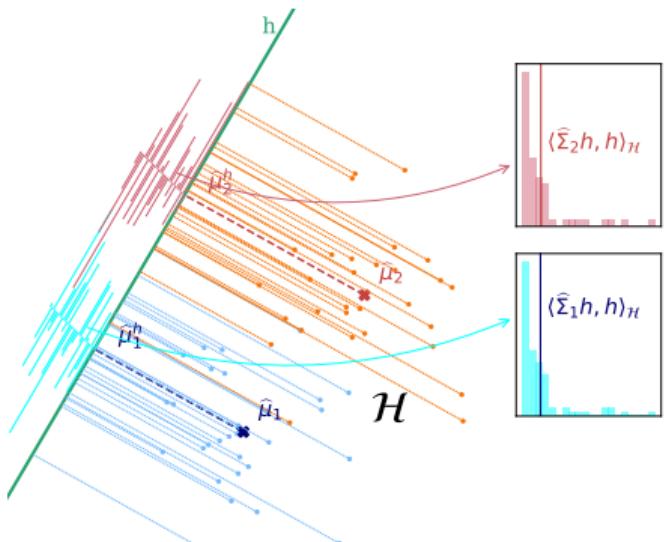
- Measure the within-group inertia of projected embeddings with

$$\langle h, \Sigma_1 h \rangle_{\mathcal{H}}, \quad \langle h, \Sigma_2 h \rangle_{\mathcal{H}}$$

- Consider an homogeneous model with averaged inertia

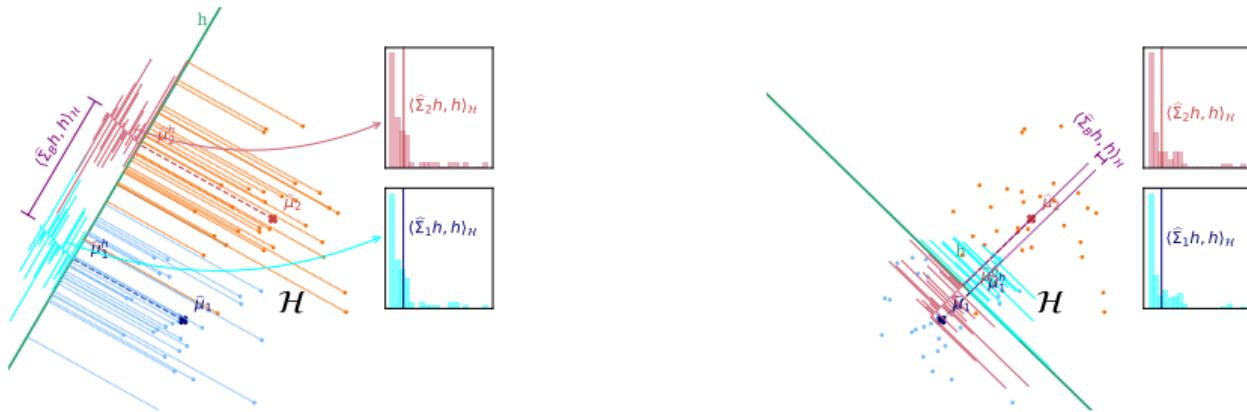
$$\langle h, \Sigma_W h \rangle_{\mathcal{H}}$$

$$\Sigma_W = \frac{n_1}{n} \Sigma_1 + \frac{n_2}{n} \Sigma_2$$



An optimisation problem

Objective: Find h such that $\langle h, \Sigma_B h \rangle_{\mathcal{H}}$ is maximal and $\langle h, \Sigma_W h \rangle_{\mathcal{H}}$ is minimal.



Define the **maximum Kernel Fisher Discriminant Ratio (KFDR)**:

$$\text{KFDR}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{h \in \mathcal{B}_{\mathcal{H}}} n \frac{\langle h, \Sigma_B h \rangle_{\mathcal{H}}}{\langle h, \Sigma_W h \rangle_{\mathcal{H}}}$$

Resolution of the KFDA problem

- Closed-form solution: the Fisher discriminant axis (witness function)

$$h^* \propto \Sigma_W^{-1}(\mu_2 - \mu_1)$$

- The KFDA statistic is kernelized Mahalanobis distance between distributions:

$$D^2(\mathbb{P}_1, \mathbb{P}_2) = \frac{n_1 n_2}{n} \left\| \Sigma_W^{-\frac{1}{2}}(\mu_2 - \mu_1) \right\|_{\mathcal{H}}^2.$$

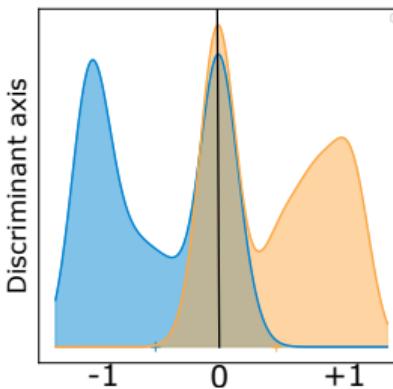
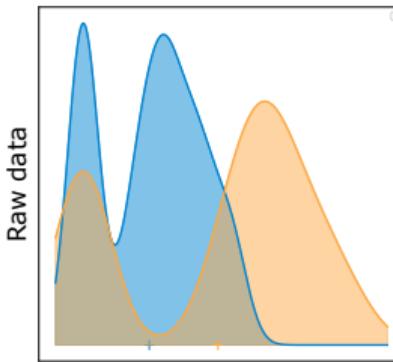
- A kernel trick is available to compute this statistic

Projection on the discriminant Axis

- Predicted labels for observations:

$$\arg \min_{i=1,2} \left\| \text{Proj} (\phi(X), h^*) - \text{Proj} (\mu_i, h^*) \right\|^2$$

- The projection onto the discriminant axis becomes a diagnostic plot for the testing procedure
- Connection with classifier-based testing



Regularized Kernel FDA testing

- The ridge strategy :

$$\Sigma_W \simeq (\Sigma_W + \gamma I_{\mathcal{H}})$$

- The spectral truncation strategy

$$\Sigma_W = \sum_{s=1}^{\infty} \lambda_s \times (f_s \otimes f_s)$$

$$\Sigma_{W,T}^{-1} = \sum_{t=1}^T \lambda_t^{-1} \times (f_t \otimes f_t)$$

- The test now depends on a hyper-parameter to tune (either γ or T)

Statistical Challenges

- The strategy consists in exploring the random variations of the test statistic under the null hypothesis $\mathbb{P}_1 = \mathbb{P}_2$.
- The target is the $(1 - \alpha)$ quantile of the distribution

$$\mathbb{P}_{H_0} \left(\widehat{\text{MMD}}^2 > q_{1-\alpha} \right) < \alpha$$

- The approximate distribution can be asymptotic / non-asymptotic
- Permutation strategies are also possible to estimate $q_{1-\alpha}$

Asymptotic distribution of the MMD [4]

- Such a threshold can be obtained through the quantiles of the asymptotic distribution of the MMD test statistic under the null hypothesis, presented in [4].

$$\widehat{nMMD}^2 \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \sum_{s \geq 1} \lambda_s \left[\left(\frac{a_s}{\sqrt{\rho_1}} - \frac{b_s}{\sqrt{\rho_2}} \right)^2 - \frac{1}{\rho_1 \rho_2} \right],$$
$$(a_s, b_s) \underset{i.i.d.}{\sim} \mathcal{N}(0, 1)$$
$$n_i/n \longrightarrow \rho_i$$

- Permutation strategies are often developed $\mathcal{O}(n^2)$.
- How to maximize power ? Aggregation is one strategy (kernel-metric learning)
- Optimality ?

Asymptotic approximations of KFDA [5, 6]

- The Ridge-regularized version

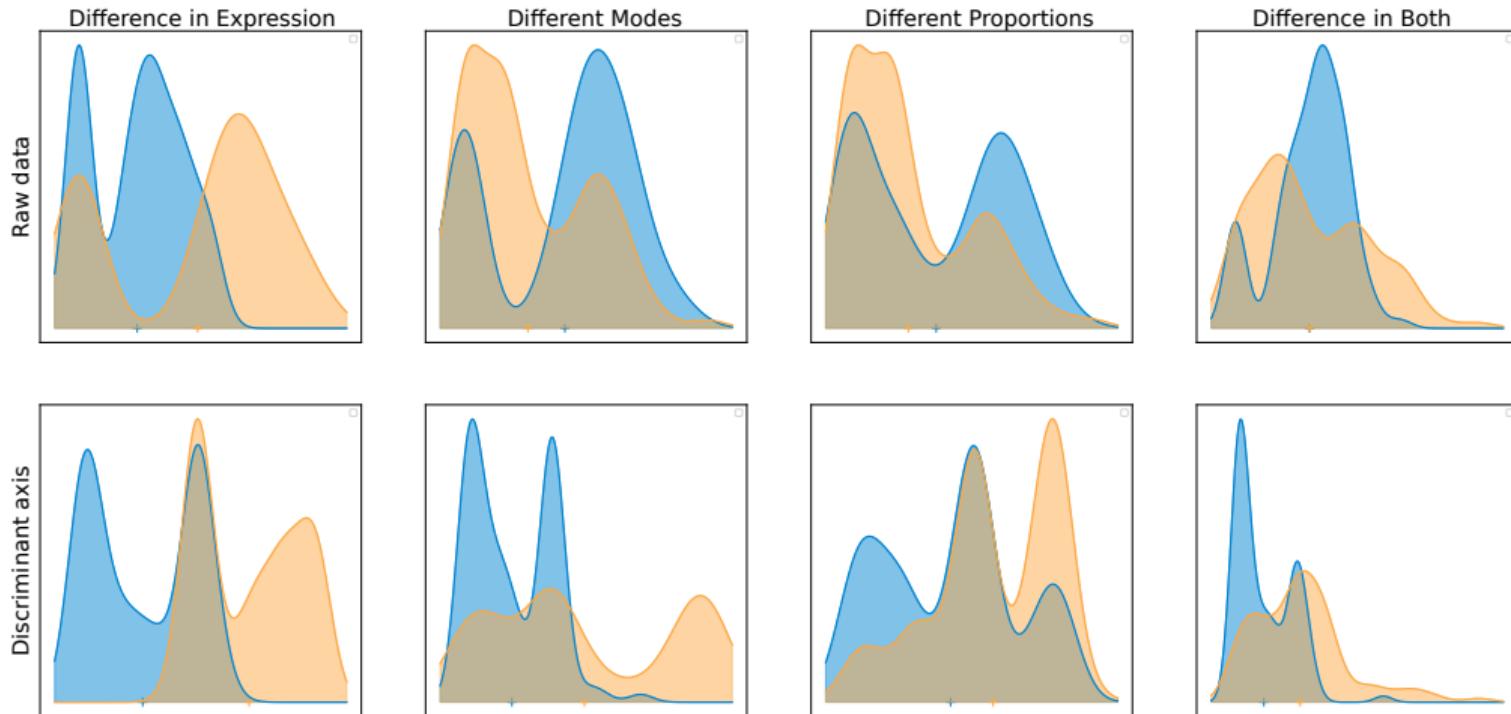
$$D_\gamma^2 \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \square(\lambda_t, \gamma) \sum_{s \geq 1} \frac{\lambda_s}{\lambda_s + \gamma} (Z_s^2 - 1), \quad Z_s \sim \mathcal{N}(0, 1)$$

- The Truncated version

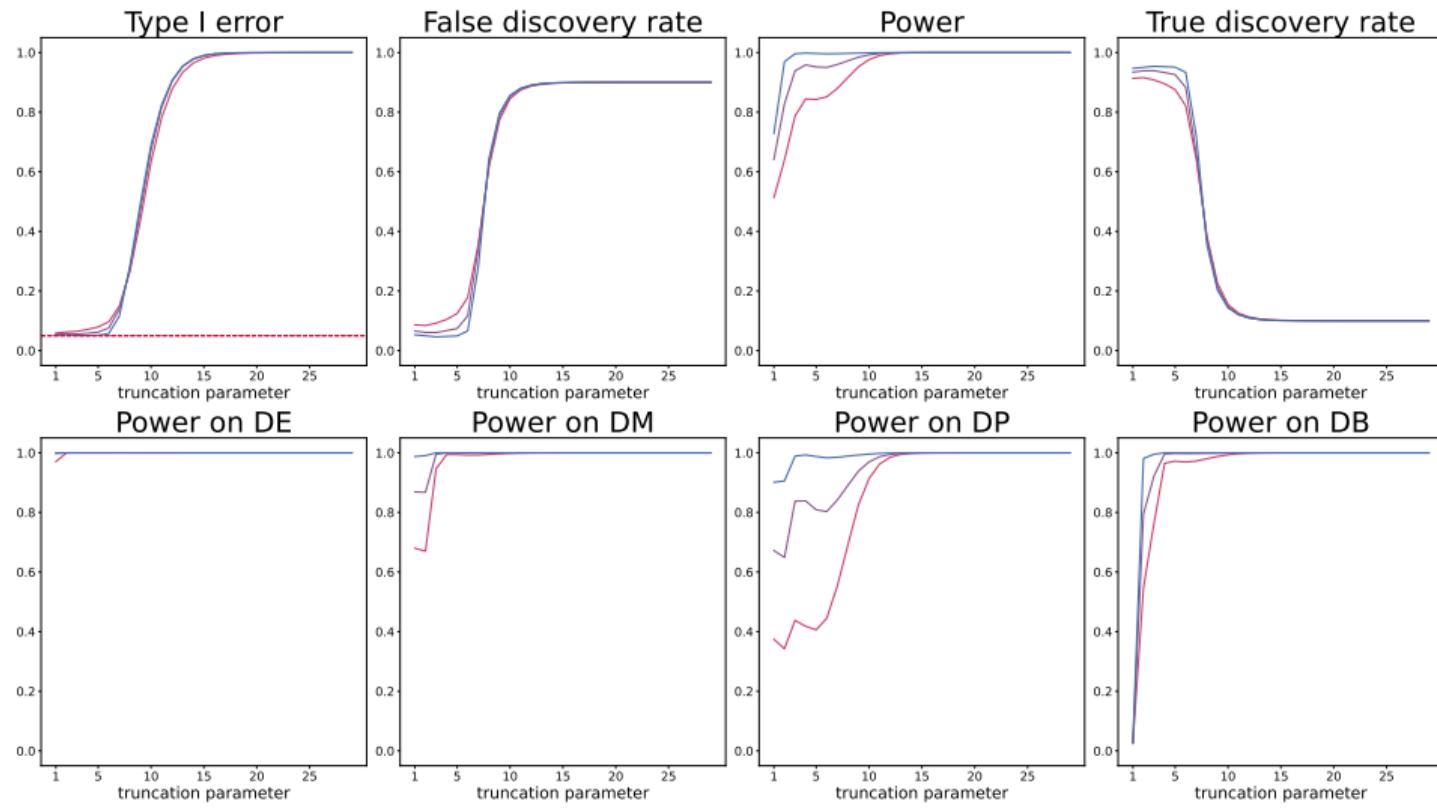
$$D_T^2 \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \sum_{t=1}^T Z_t^2 \sim \chi^2(T), \quad Z_t \sim \mathcal{N}(0, 1)$$

- The calibration issue of T

Simulated distributions



Influence of the truncation parameter

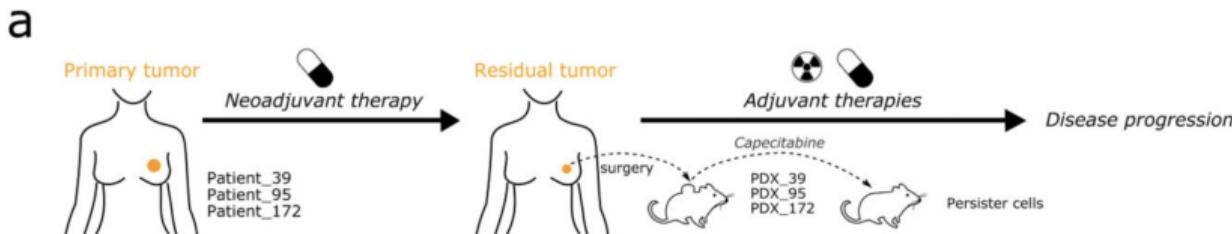
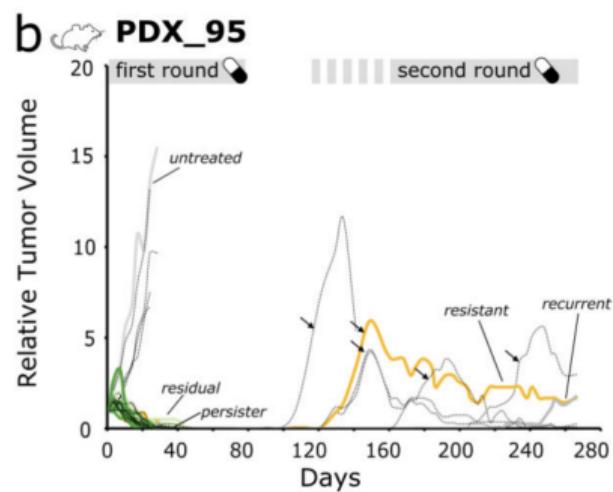


Outline

1. The Single-Cell Revolution
2. Introduction to kernel mean embedding
3. Introduction to kernel testing
4. ChemoResistance of Breast Cancer-Cells
5. Too many perspectives !

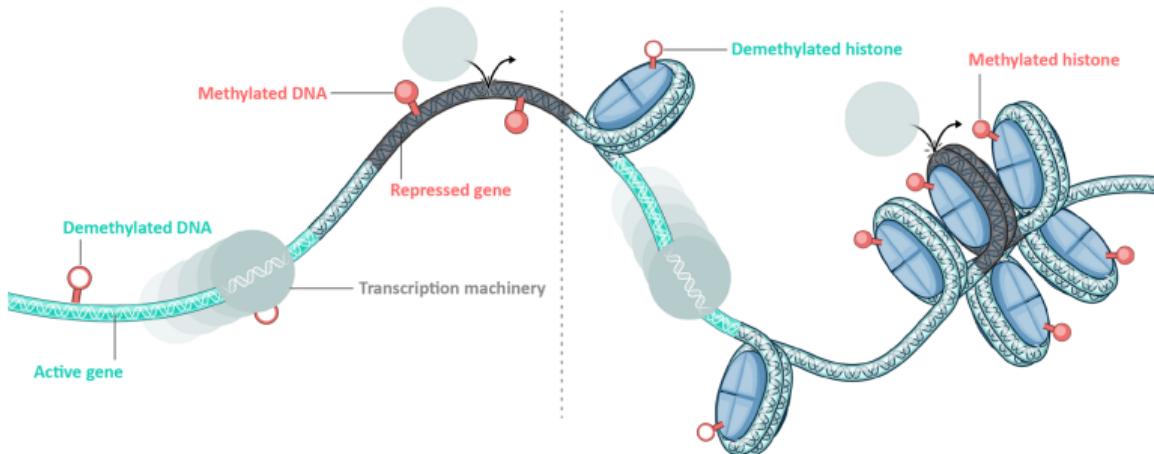
ChemoResistance in Triple Negative Breast Cancer

- Emergence of resistant phenotypes is a multi-step process
- After drug insult only a pool of drug-tolerant persister cells manage to tolerate the treatment and survive.
- Reservoir from which drug-resistant cells can ultimately emerge.



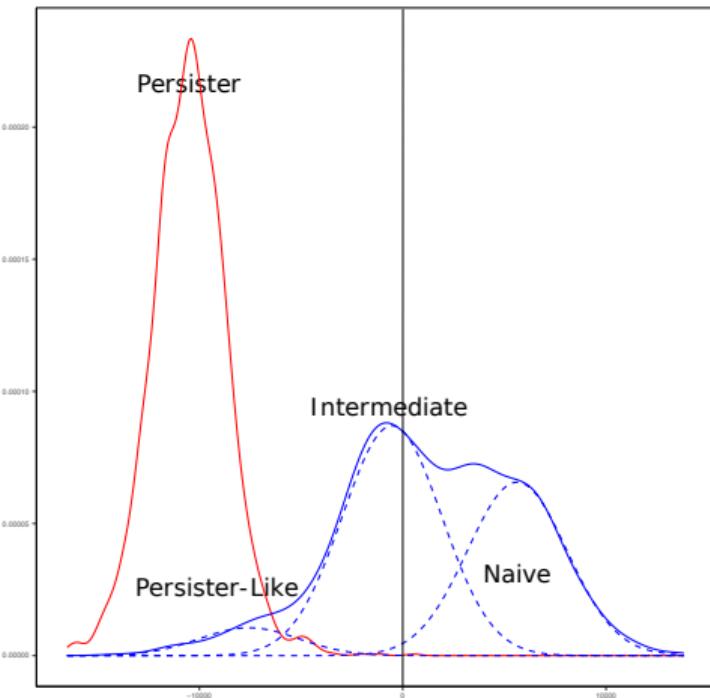
Epigenomics

- Chromatin modifications are a key actor of gene expression regulations
- Variations in chromatin modifications induce plasticity
- ChemoResistance is associated with epigenomic variations



Kernel testing on Persister vs. Naive cells

- Persister cells survived the first treatment
- Reservoir for resistant cells
- Epigenomic data: 6376 features
- Compare untreated (~ 3000 cells) vs. persister (~ 2000 cells)
- Project distributions on the discriminant axis
- Did we identify the reservoir of persister cells based on their epigenomic signatures ?



Check out !

- The ktest package ! Python-R, fisrt package on kernel testing (including MMD, KFDA, Nystrom)

<https://github.com/AnthoOzier/ktest>

- The arxiv preprint

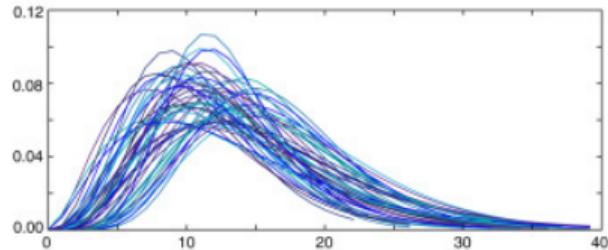
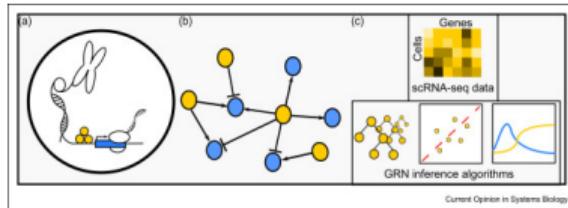
<https://arxiv.org/abs/2307.08509>

Outline

1. The Single-Cell Revolution
2. Introduction to kernel mean embedding
3. Introduction to kernel testing
4. ChemoResistance of Breast Cancer-Cells
5. Too many perspectives !

How to reconcile global and feature-wise testing ?

- Kernel testing performs global testing of joint multivariate distributions based on $\mathbf{X}_i = [X_i^1, \dots, X_i^G]$
- KFDA is a powerful metric to compare joint distributions
- How to determine which features support most the rejection ?
- Interpretability and sensitivity analysis
(D. Garreau, U. Nice)



Distribution of gene expression across cells

Generalize the two-sample framework !

- The two-sample framework corresponds to a one-way ANOVA on embeddings

$$\phi(X_{i,j})(\bullet) = \mu(\bullet) + \alpha_i(\bullet) + \varepsilon_{i,j}(\bullet)$$

- Kernel-FDA is a test of contrast : $H_0 : \{\alpha_1 = \alpha_2\}$.
- Generalize to the general setting of the linear model :

$$\phi(\mathbf{X})(\bullet) = \mathbf{W}\boldsymbol{\theta}(\bullet) + \mathbf{E}(\bullet)$$

- Test linear functional hypothesis $\mathbf{C}\boldsymbol{\theta} = 0$
- Develop diagnostic plots for kernel testing (Cook, influence)
- See Anthony's Thesis !

How good is the asymptotics ?

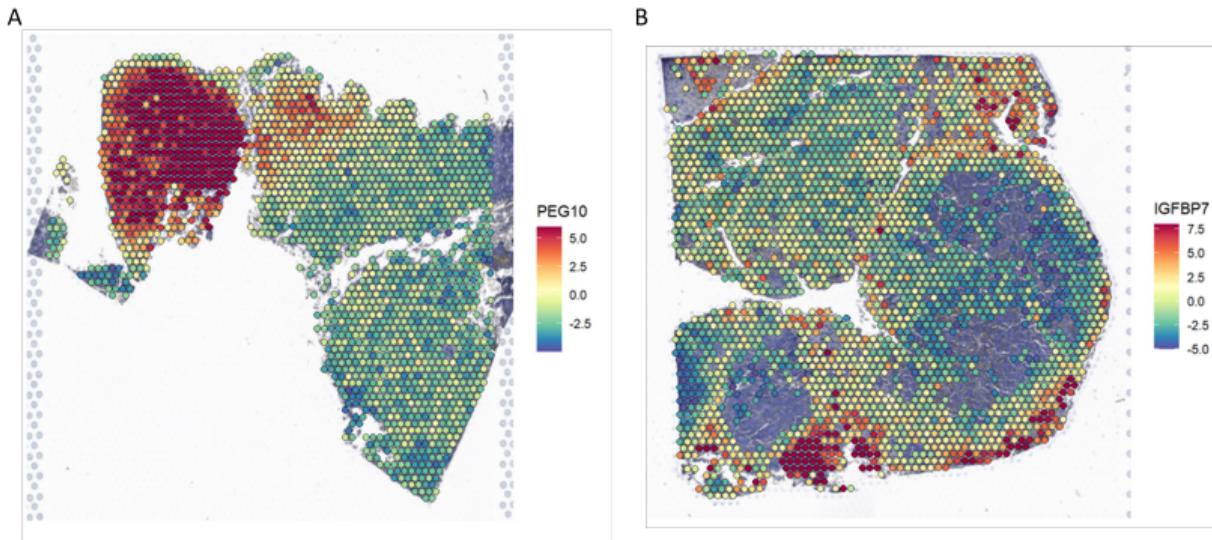
- The procedure is based on an asymptotic approximation :

$$D_T^2 \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \sum_{t=1}^T Z_t^2 \sim \chi^2(T), \quad Z_t \sim \mathcal{N}(0, 1)$$

- Could it be non-asymptotic ? (postdoc Perrine Lacroix, with V. Rivoirard, B. Michel)
- What about the theoretical calibration (T)?
- Learn the kernel ?
- In practice, too many rejections ! How to relax the null hypothesis ?

Next Challenge ahead: Spatial Transcriptomics

- Spatial Transcriptomics consists in measuring gene expression directly on the tissue
- Each cell has a spatial coordinates
- Adapt Kernel methods to spatialized data (spatial HSIC ?)



→ Positions available !!!

References

- [1] S. Aldridge and S. A. Teichmann. Single cell transcriptomics comes of age. *Nat Commun*, 11(1):4307, Aug 2020.
- [2] J. N. Campbell, E. Z. Macosko, H. Fenselau, T. H. Pers, A. Lyubetskaya, D. Tenen, M. Goldman, A. M. Verstegen, J. M. Resch, S. A. McCarroll, E. D. Rosen, B. B. Lowell, and L. T. Tsai. A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.*, 20(3):484–496, Mar 2017.
- [3] J. Fan, K. Slowikowski, and F. Zhang. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp Mol Med*, 52(9):1452–1465, Sep 2020.
- [4] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A Kernel Two-Sample Test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.
- [5] Z. Harchaoui, F. Bach, and E. Moulines. Testing for Homogeneity with Kernel Fisher Discriminant Analysis. *arXiv:0804.1026 [stat]*, Apr. 2008.
- [6] Z. Harchaoui, F. Vallet, A. Lung-Yut-Fong, and O. Cappe. A regularized kernel-based approach to unsupervised audio segmentation. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1665–1668, Taipei, Taiwan, Apr. 2009. IEEE.
- [7] K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf. 2017.
- [8] A. Regev, S. A. Teichmann, E. S. Lander, and I. e. a. Amit. The Human Cell Atlas. *Elife*, 6, 12 2017.
- [9] H. van Assel, T. Espinasse, J. Chiquet, and F. Picard. A probabilistic graph coupling view of dimension reduction. *Arxiv*, (2201.13053):1–15, 2022.