

Application en BDR de l'Intelligence Artificielle : quelques exemples

Guillaume BACHELOT

Service de biologie de la reproduction – CECOS

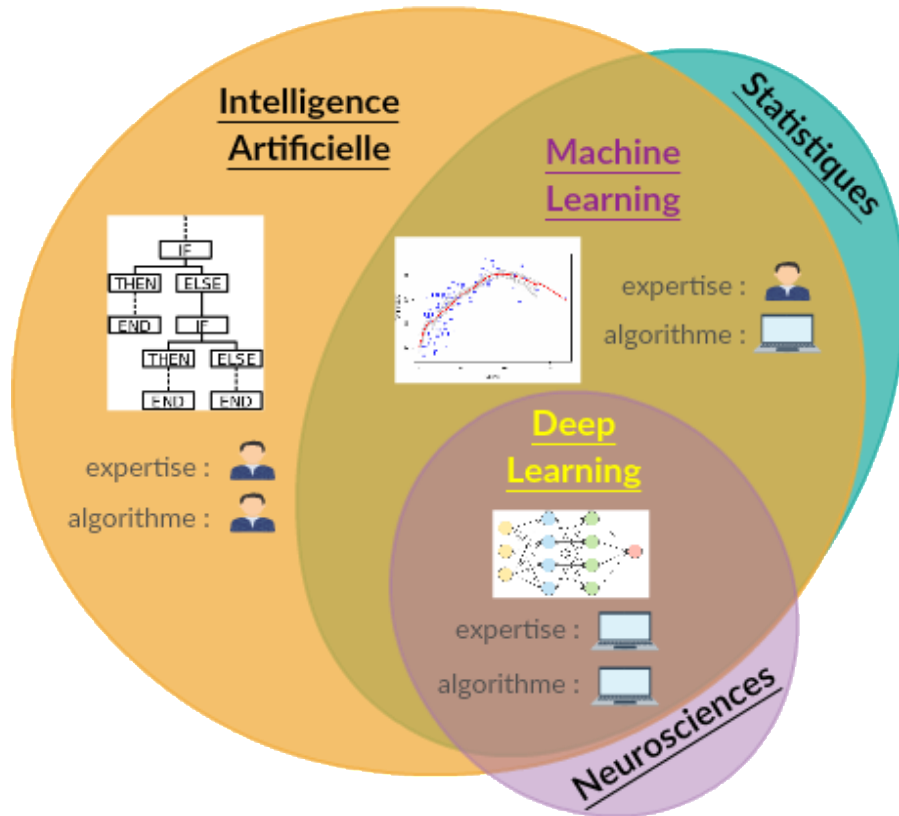
Hôpital Tenon, APHP.Sorbonne Université



Plan

- Introduction
- Exemple 1 : Prédiction de l'extraction de spermatozoïdes par biopsie testiculaire
- Exemple 2 : Identification des hyperplasies congénitales des surrénales et diagnostic différentiel avec le SOPK
- Conclusion

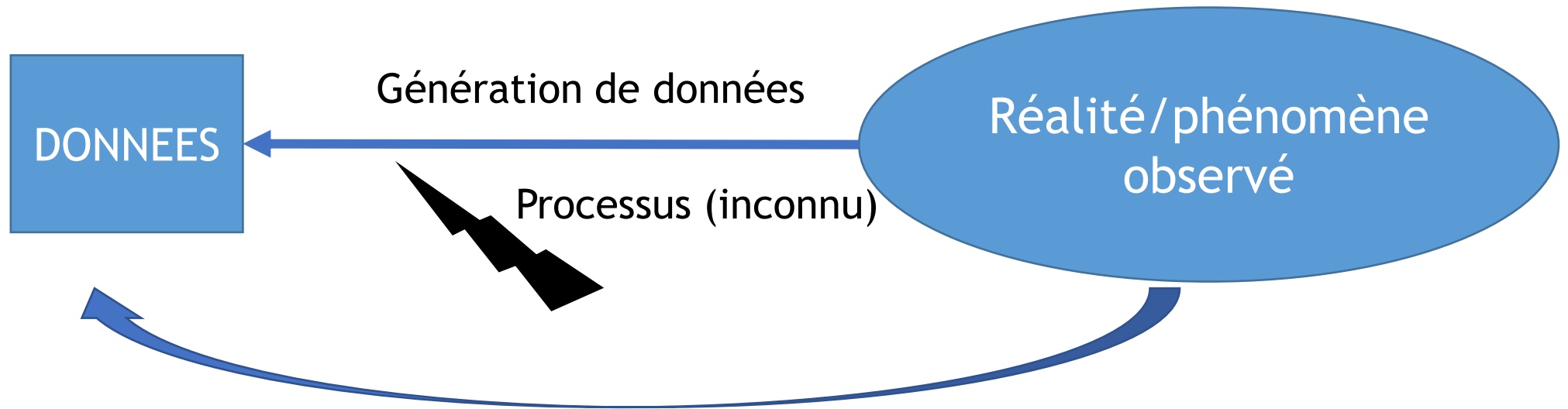
Introduction



La **modélisation statistique** est une manière simplifiée et formalisée mathématiquement de s'approcher de la réalité et, en d'autres termes, de décrire les processus qui génèrent vos données.

Elle permet de faire des prédictions à partir de cette approximation. Le modèle statistique est l'équation mathématique utilisée.

Introduction



Modéliser et estimer relation entre phénomène \leftrightarrow données
Pour passer de l'un à l'autre
De manière automatique \rightarrow par l'exemple !

Introduction

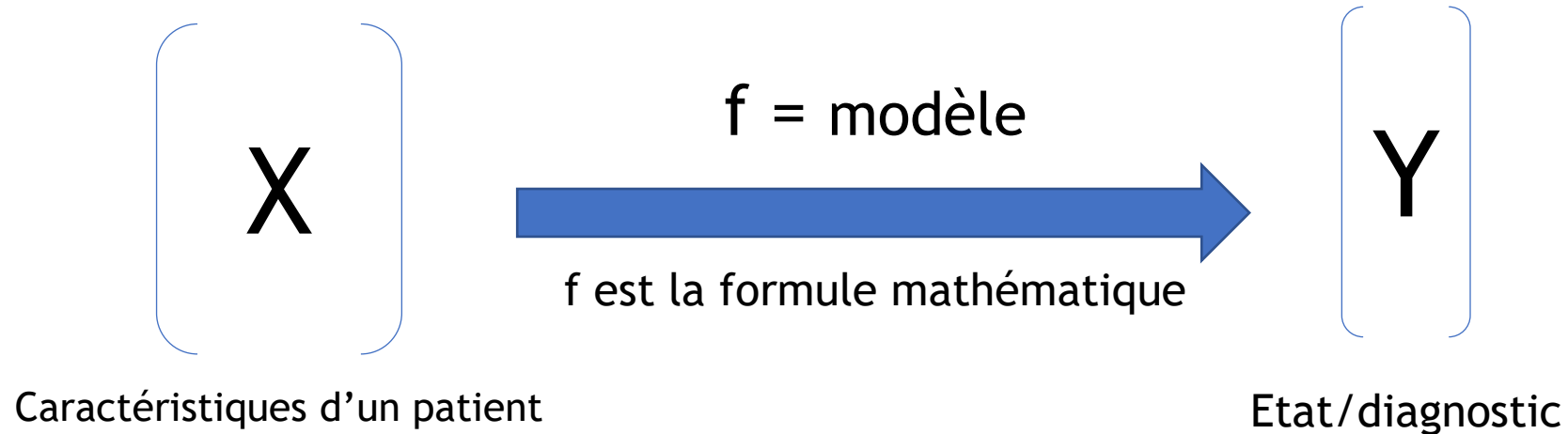
Biopsie testiculaire

SOPK / CAH

Conclusion

Introduction

→ Apprendre la fonction $f(X) = Y...$

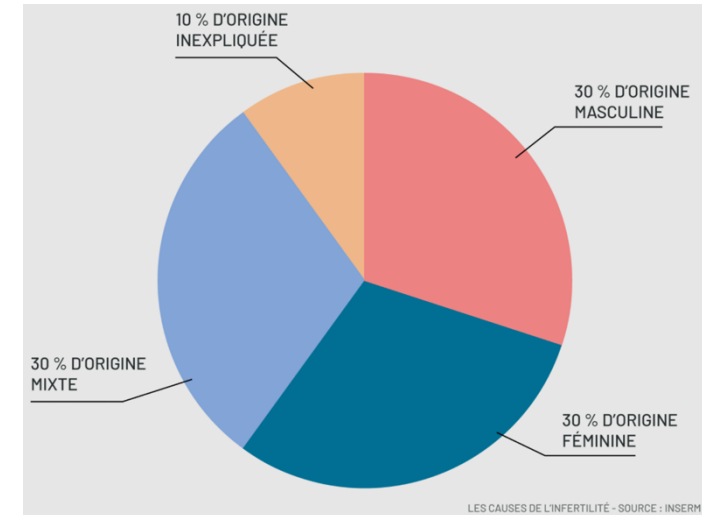
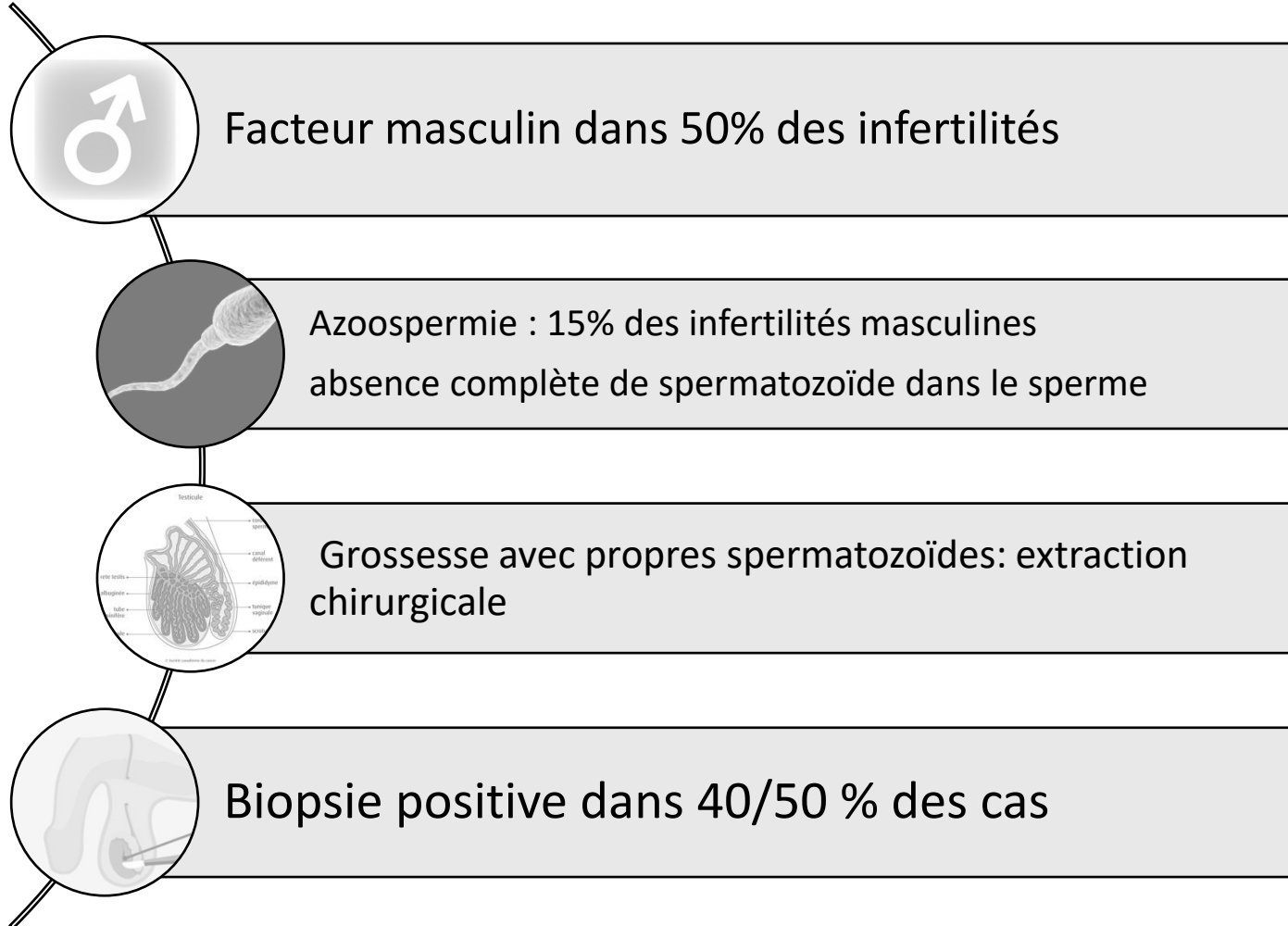


...de manière automatique, en apportant
des exemples connus (couples X - Y)

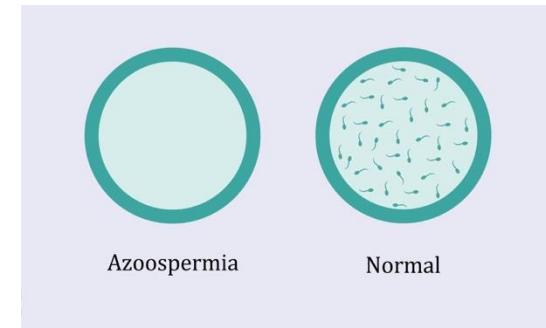
Objectif : Prédire ou identifier Y à partir de X pour de nouveaux patients

Exemple 1 : Prédiction de l'extraction de spermatozoïdes par biopsie testiculaire

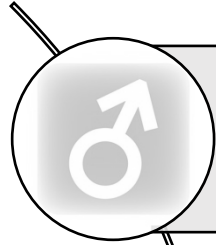
Contexte de la biopsie testiculaire et cas d'usage



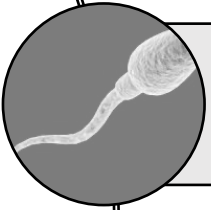
± 15 %



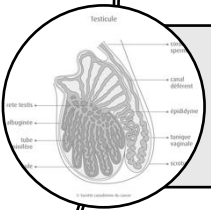
Contexte de la biopsie testiculaire et cas d'usage



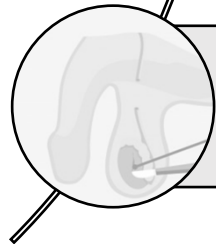
Facteur masculin dans 50% des infertilités



Azoospermie : 15% des infertilités masculines
absence complète de spermatozoïde dans le sperme

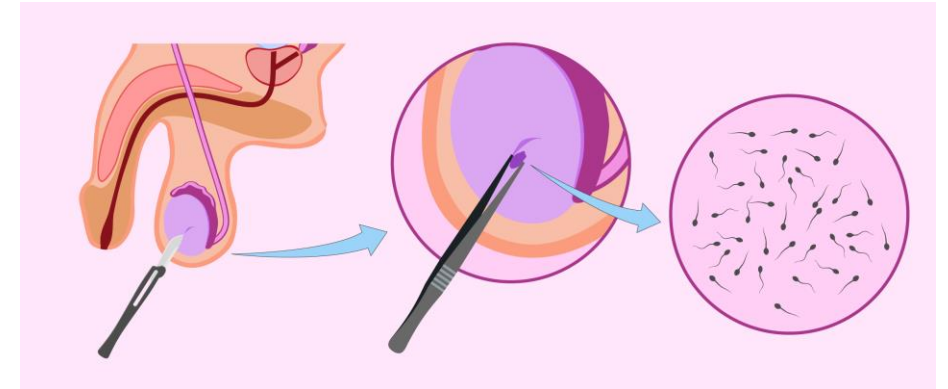


Grossesse avec propres spermatozoïdes: extraction chirurgicale

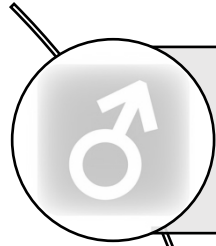


Biopsie positive dans 40/50 % des cas

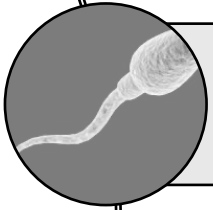
± 50 % de chance de
récupérer des spermatozoïdes



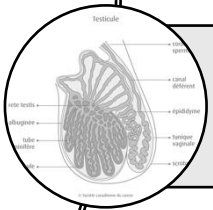
Contexte de la biopsie testiculaire et cas d'usage



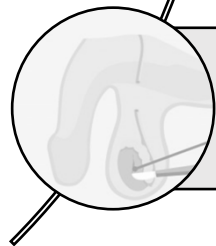
Facteur masculin dans 50% des infertilités



Azoospermie : 15% des infertilités masculines
absence complète de spermatozoïde dans le sperme



Grossesse avec propres spermatozoïdes: extraction chirurgicale



Biopsie positive dans 40/50 % des cas

Problématique

Peu d'outils - facteurs prédictifs
fiables chances de succès
Évaluation balance bénéfice/risque

Hypothèse

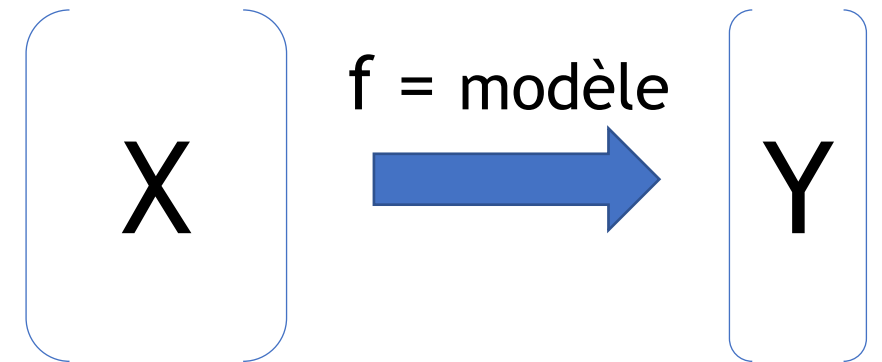
Combinant bilan préopératoire +
apprentissage automatique → Score
prédictif du résultat de la biopsie
testiculaire

Cas d'usage de l'apprentissage automatique

→ Apprendre la fonction $f(X) = Y...$

→ Modélisation la relation entre la présence de spermatozoïdes ou non à la biopsie testiculaire (= phénomène) et les informations du bilan préopératoire

→ Objectif : pouvoir prédire résultat de la biopsie à partir du bilan préopératoire

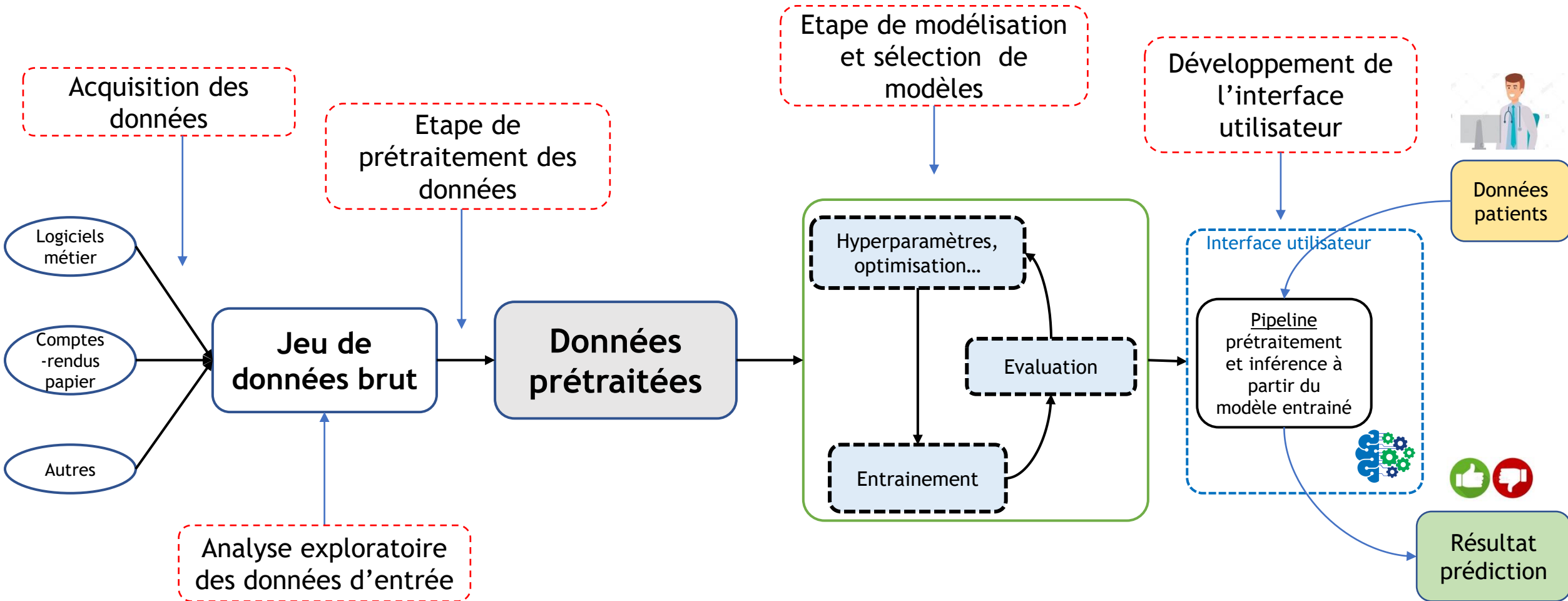


Caractéristiques d'un patient
= Bilan préopératoire

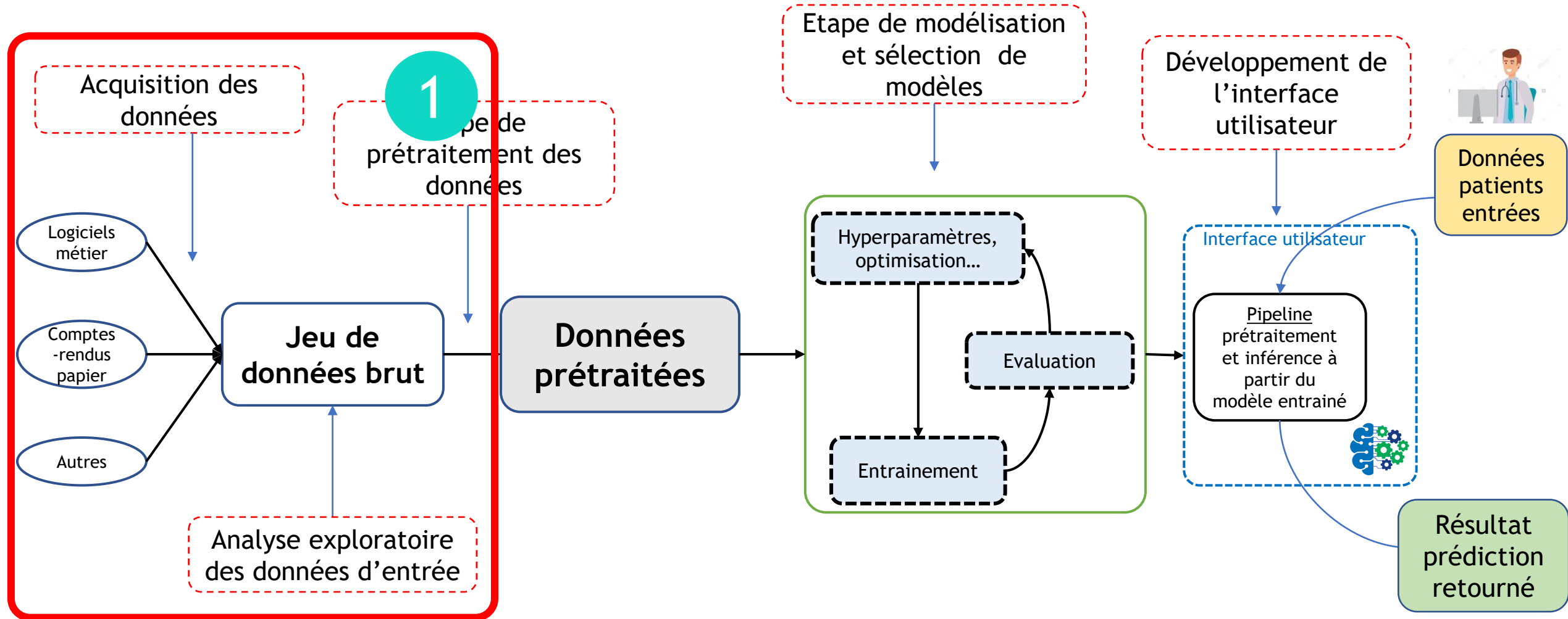
Etat/diagnostic:
= Biopsie + ou -

...de manière automatique, en apportant des exemples (couples X - Y)

En pratique ? Les étapes de développement



1. Acquisition et analyse exploratoire des données



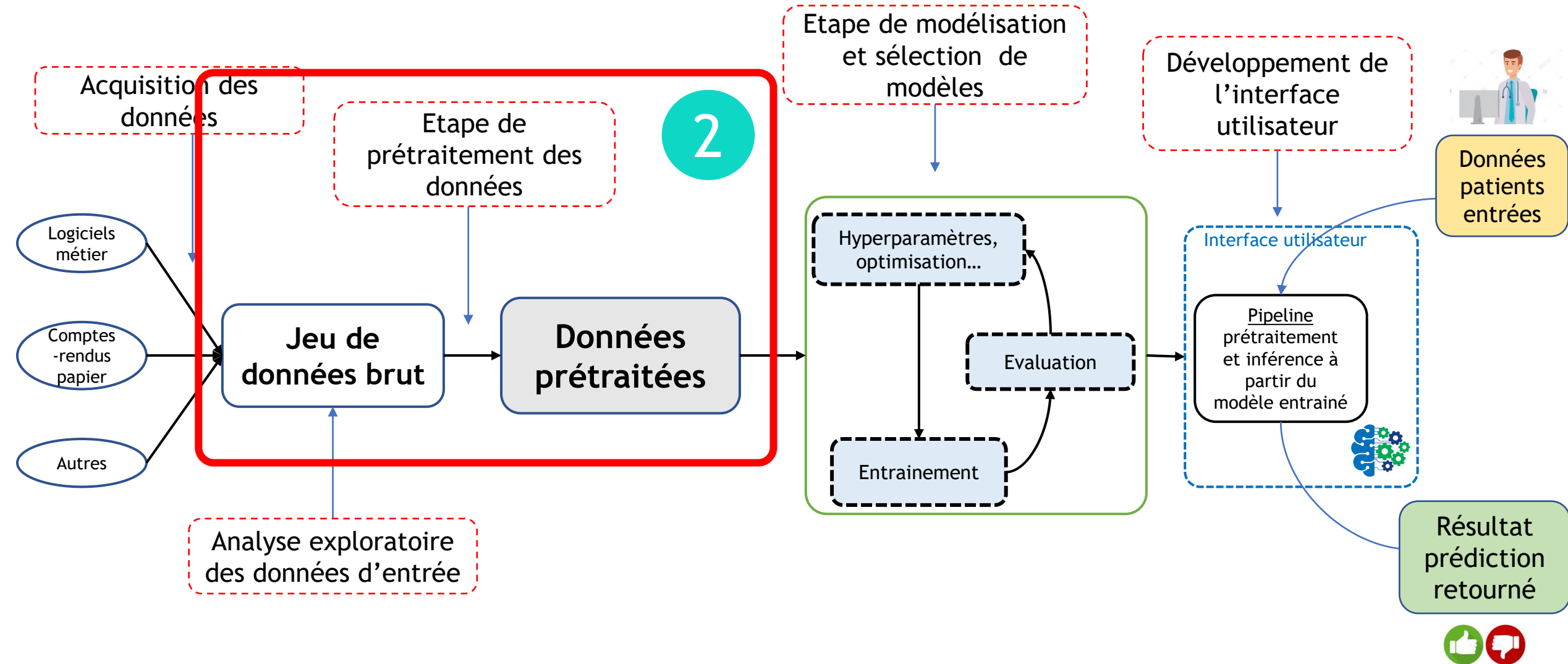
Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

2. Prétraitement des données



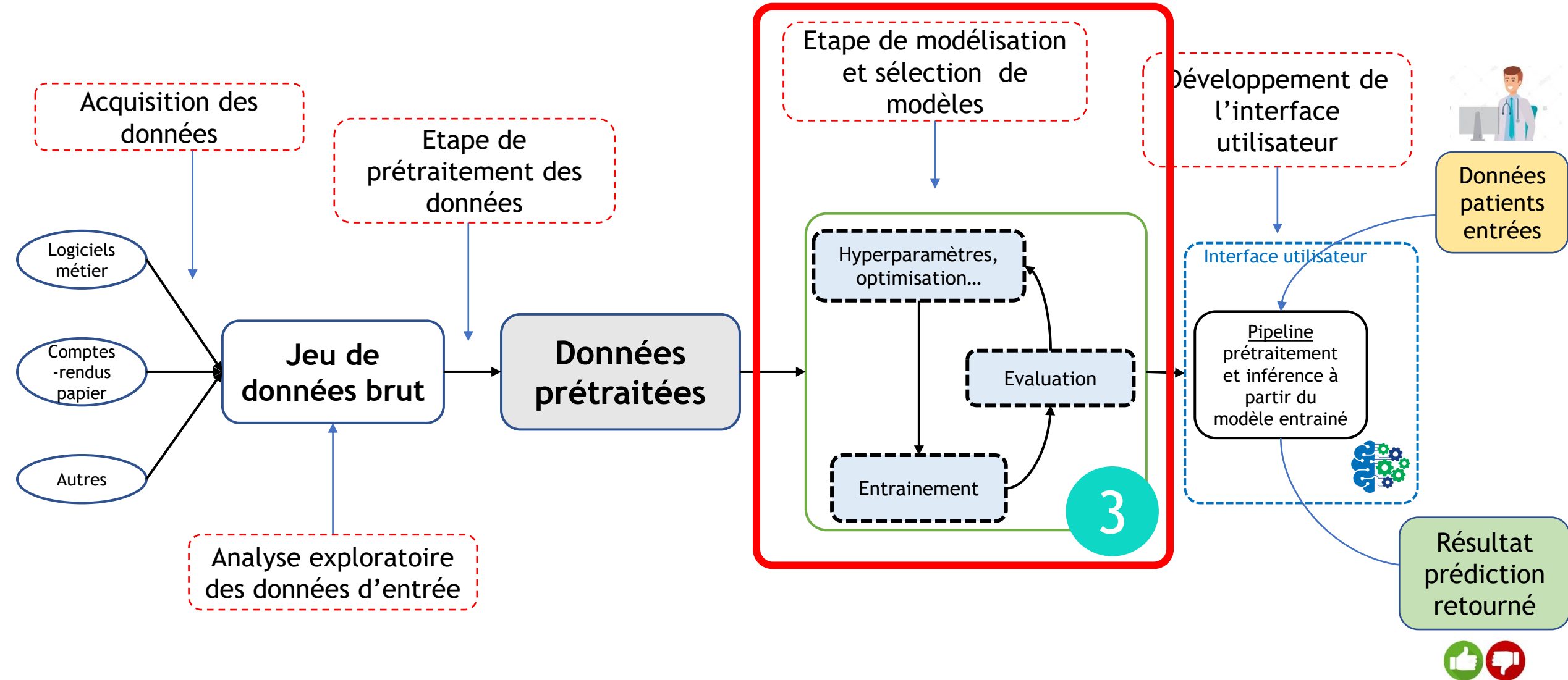
Introduction

Biopsie testiculaire

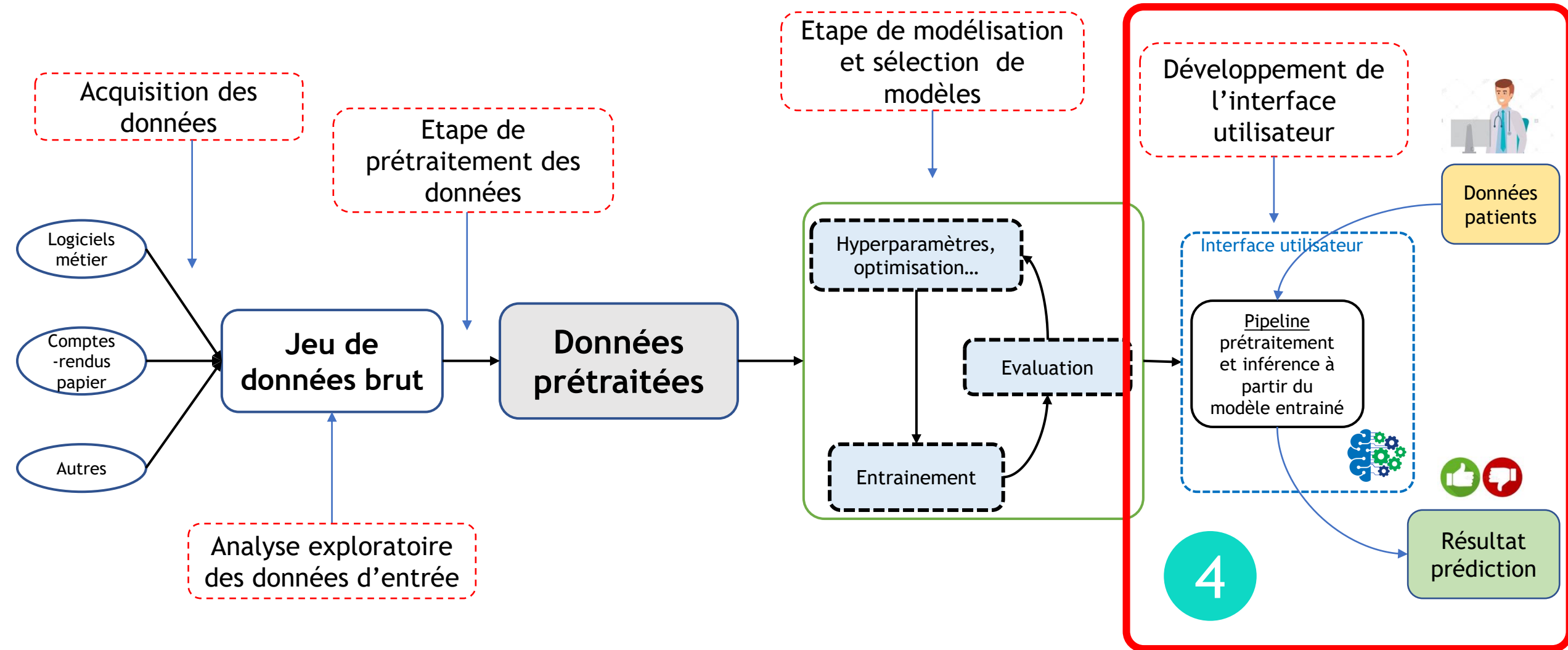
SOPK / CAH

Conclusion

3. Modélisation : processus itératif



4. Application web pour utilisateur final



De quoi a-t-on besoin ? patients, données et modèles

Données (= exemples)

- n = 201 Patients (azoospermie NO)
- Mars 2014 – décembre 2021 Hôpital Tenon
- Bilans préop' (X) + résultat BT (Y)

Modèles utilisés et comparés

• Machine Learning

→ classification naïve bayésienne, régression logistique, méthode des k plus proches voisins, machine à vecteurs de support, méthodes ensemblistes : forêt aléatoire et dérivés (*GradientBoosting*, *XGBoost*...)

• Deep Learning

→ plusieurs architectures

	Variable	Manquants (%)	Biopsie positive (n=104)	Biopsie négative (n=71)	p-value
Clinique	Âge (années)	14,3	38.11 (9.56)	36.89 (5.40)	NS
	BMI	12,4	25.41 (4.99)	25.11 (3.65)	NS
	Tabac	10,3	27/94 (28,7%)	19/63 (30,2%)	NS
Hormonologie	FSH (mmol/L)	9,7	9.90 (14.10)	23.50 (23.03)	<0.001
	LH (mmol/L)	21,1	6.37 (5.49)	11.90 (12.80)	<0.001
	Testostérone (mmol/L)	21,1	4.50 (2.76)	4.00 (3.05)	NS
	Inhibine B (mmol/L)	22,1	60.00 (94.00)	14.00 (28.25)	<0.001
	Prolactine (mmol/L)	31,4	12.00 (5.58)	11.85 (8.08)	NS
Génétique	Caryotype normal	8,0	9/87 (10,3%)	6/64 (9,4%)	NS
	Microdélétion Y	8,6	2/96 (2,1%)	3/64 (4,7%)	NS
Antécédents	Cryptorchidie	5,7	20/98 (20,4%)	11/67 (16,4%)	NS
	Infection	7,4	11/97 (11,3%)	3/65 (4,6%)	NS
	Traumatisme	6,9	2/97 (2,1%)	1/66 (1,5%)	NS
	Traitement gonadotoxique	6,9	8/98 (8,2%)	7/65 (10,8%)	NS
	Chirurgie uro-génitale	7,4	11/97 (11,3%)	6/65 (9,2%)	NS
	Varicocèle	5,7	14/99 (14,1%)	15/66 (22,7%)	NS

Data are n/N (%), or median (IQR).

Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

De quoi a-t-on besoin ? patients, données et modèles

Données (= exemples)

- n = 201 Patients (azoospermie NO)
- Mars 2014 – décembre 2021 Hôpital Tenon
- Bilans préop' (X) + résultat BT (Y)

Modèles utilisés et comparés

• Machine Learning

→ classification naïve bayésienne, régression logistique, méthode des k plus proches voisins, machine à vecteurs de support, méthodes ensemblistes : forêt aléatoire et dérivés (*GradientBoosting*, *XGBoost*...)

• Deep Learning

→ plusieurs architectures

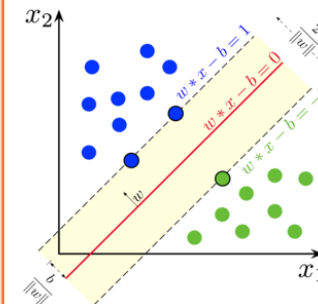
Bilan préopératoire

BT + ou - ?

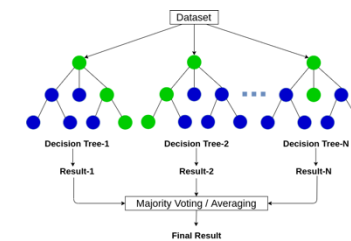
X

Fonction

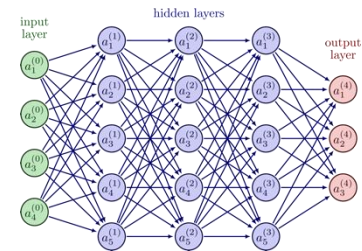
Y



SVM



Random forest



Réseau neurones

Introduction

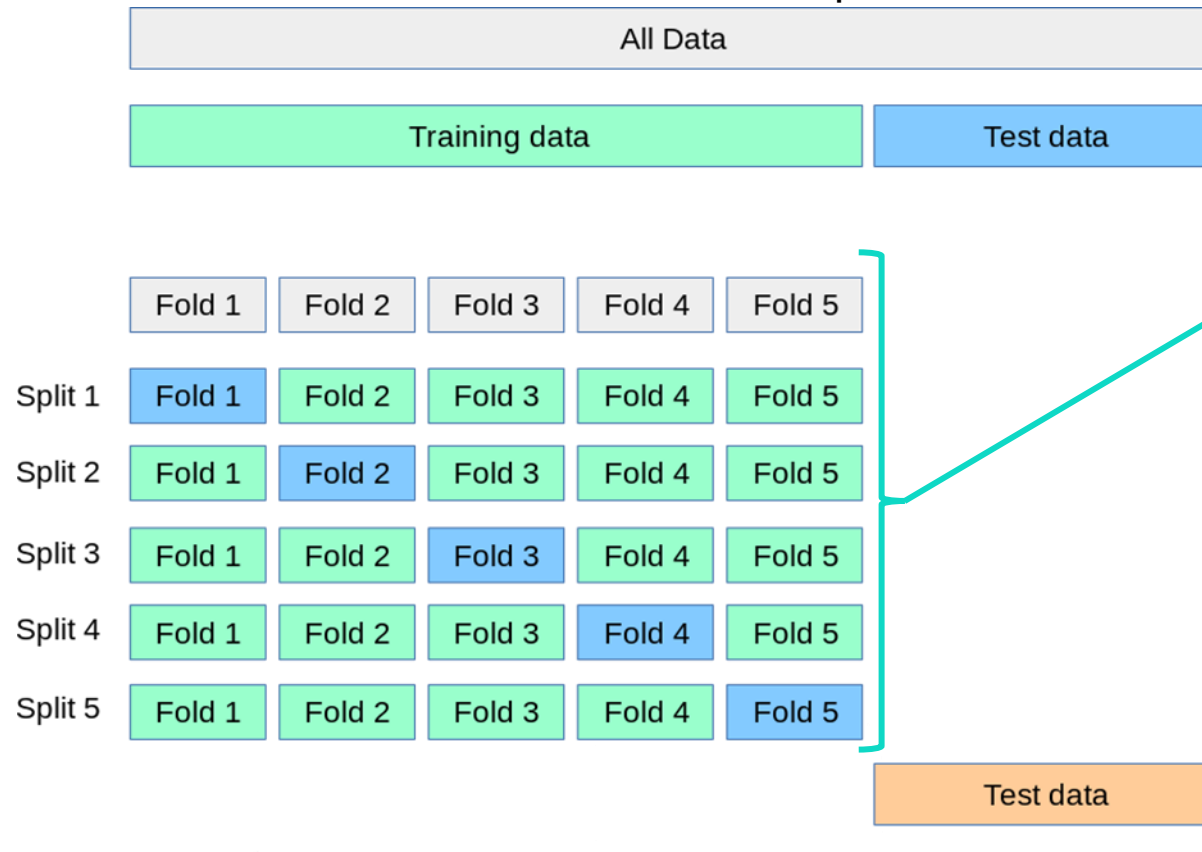
Biopsie testiculaire

SOPK / CAH

Conclusion

Comment mesurer performance du modèle construit ? Stratégie d'évaluation

Etude monocentrique



Cohorte d'entraînement

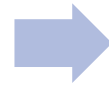
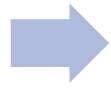
- Entraîner et régler le modèle
- Recherche hyperparamètres
- Évaluation interne par validation croisée

Cohorte de test

- Évaluation finale « externe »

Métriques d'évaluation ?

- Sensibilité
- Spécificité
- Exactitude
- AUC ROC



Résultats : comment sont-ils présentés ?

- Q principale : Modèles les plus performants ?
 - Q secondaire (1) : Variables les plus pertinentes ?
 - Q secondaire (2) : Taille d'échantillon optimale ?

Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

Modèles les plus performants

Cohorte totale = 201 patients

Cohorte d'entraînement = 175 patients

Cohorte de test =
26 patients

Model	Internal validation (cross validation)				External temporal validation			
	AUC ROC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC ROC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
LR	66,7	68,6	79,8	52,1	85,2	69,2	76,9	61,5
BNC	63,0	65,1	71,2	56,3	83,4	69,2	69,2	69,2
RF	77,4	74,9	85,6	59,2	89,6	84,6	100,0	69,2
GBT	79,4	73,7	79,8	62,0	82,2	76,9	92,3	61,5
XGB	77,3	72,6	78,8	55,9	81,7	80,8	92,3	69,2
SVM	71,2	68,6	82,7	47,9	72,2	61,5	84,6	38,5
KNN	65,3	64,6	76,0	47,9	76,3	69,2	84,6	53,8
ANN (64x32)	70,7	64,6	66,4	61,9	65,4	65,0	54,0	77,0

Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

Les résultats d'un modèle dans le détail

Cohorte totale = 201 patients

Cohorte d'entraînement = 175 patients

Cohorte de test = 26 patients

Modèle	Validation interne (validation croisée)					Validation externe				
	AUC ROC (%)	Exactitude (%)	Sensibilité (%)	Spécificité (%)		AUC ROC (%)	Exactitude (%)	Sensibilité (%)	Spécificité (%)	
Random Forest	80,6	78,3	83,7	70,4		91,9	83,9	84,2	83,3	
Gradient boosting	62,4	68,8	75,2	45,2		66,3	68,7	77,1	41,7	
Classification naïve bayésienne	66,6	69,1	79,8	53,5		65	71	79	41,7	
Régression logistique	66,6	68	81,7	47,9		79,4	71	89,5	41,7	
Support Vector Classification	64,5	65,1	76,9	47,9		8	74,2	89,5	50	
K-nearest neighbor	80,6	78,3	83,7	70,4		91,9	83,9	84,2	83,3	
	77,1	81,2	83,7	70,4		85,5	77,4	73,7	83,3	
K-nearest neighbor	60,6	68,8	75,2	45,2		61,1	67,7	84,2	41,7	
	75,8	80,6	89,5	66,7		90,8	80,6	89,5	66,7	
	72,3	80,6	89,5	66,7		87	80,6	89,5	66,7	
	68,6	80,6	89,5	66,7		87	80,6	89,5	66,7	

Moyenne des résultats des 5
« blocs » du découpage de
validation croisée

Performances sur la cohorte
de test qui sont les plus
représentatifs des
performances FUTURES de
l'algorithme !!

Journée des Techniciens et Biologistes en AMP

Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

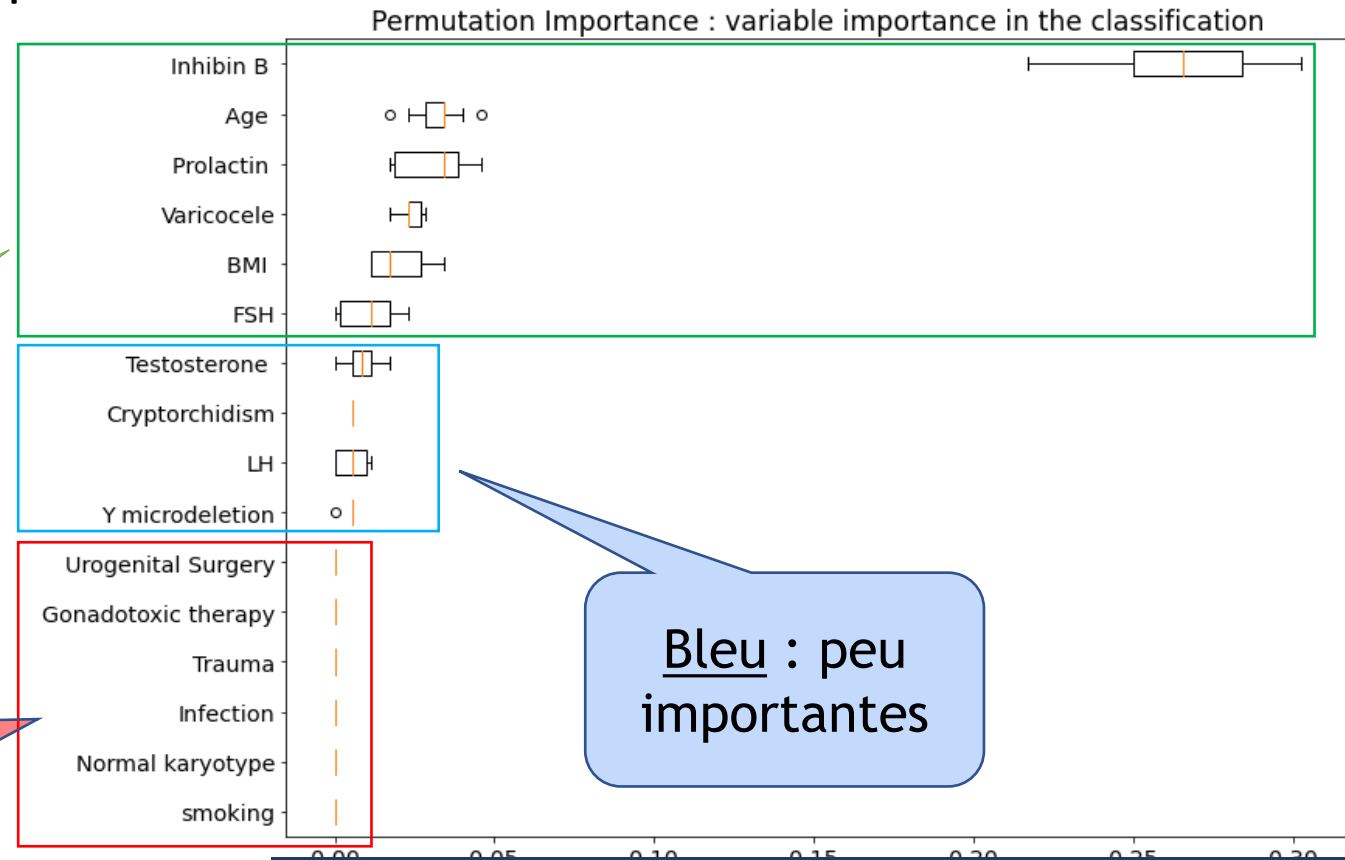
Les variables les plus pertinentes

- Les variables peuvent être classées selon leur importance dans le modèle, selon leur pouvoir discriminant/informatif

Vert :
variables
importantes

Rouge :
n'apportent aucune
information voire
du bruit

Bleu : peu
importantes



Taille d'échantillon optimale

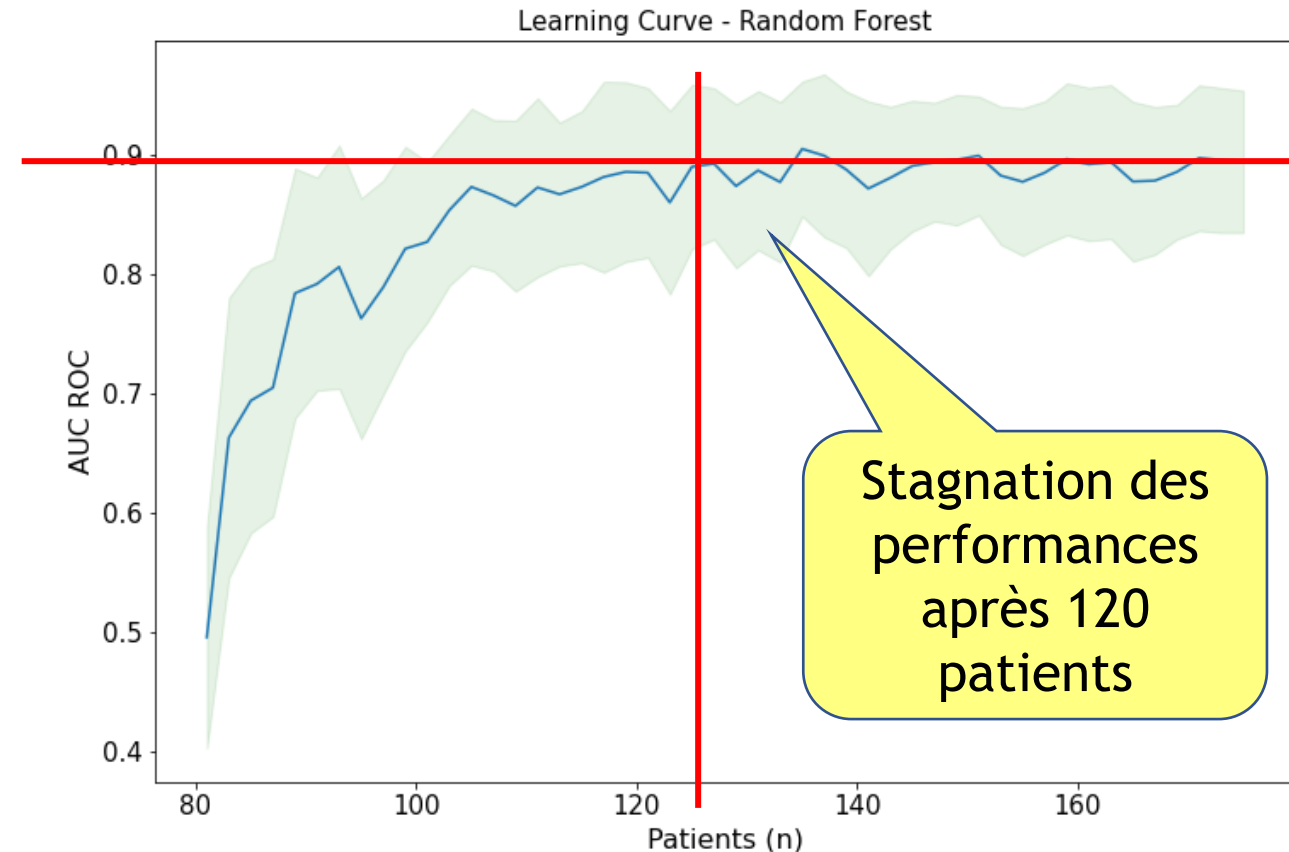
- Courbe d'apprentissage

- ➔ Evolution des performances du modèles (ordonnée) au fur et à mesure ajout patients dans cohorte d'entraînement (abscisse)

- ➔ Allure de la courbe ?

- Augmentation des performances jusqu'à 120 / 130 patients, puis plateau !

- Intérêt d'une inclusion illimité ?

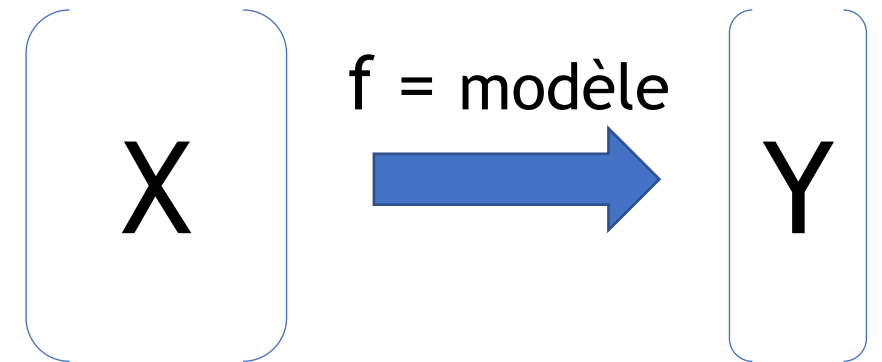


Conclusion exemple 1 → quel regard critique ?

Rappel :

- Modélisation la relation entre la présence de spermatozoïdes ou non à la biopsie testiculaire (= phénomène) et les informations du bilan préopératoire
- Objectif : pouvoir prédire résultat de la biopsie à partir du bilan préopératoire

→ Apprendre la fonction $f(X) = Y...$



Caractéristiques d'un patient
= Bilan préopératoire

Etat/diagnostic:
= Biopsie + ou -

...de manière automatique, en apportant des exemples (couples X - Y)

Conclusion exemple 1 → quel regard critique ?

1. Sur les exemples choisis : cohorte représentative ?

- Données rétrospectives, 1 seul centre

→ Apprendre la fonction $f(X) = Y...$

- Validation multicentrique ou prospective nécessaire

2. → Résultats intéressants (modèle > variables univariées)

3. → Limites +++ utilisation données rétrospectives / non conçues/récoltées dans cet objectif

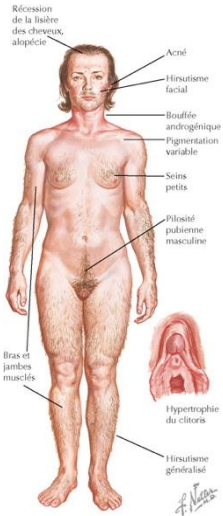
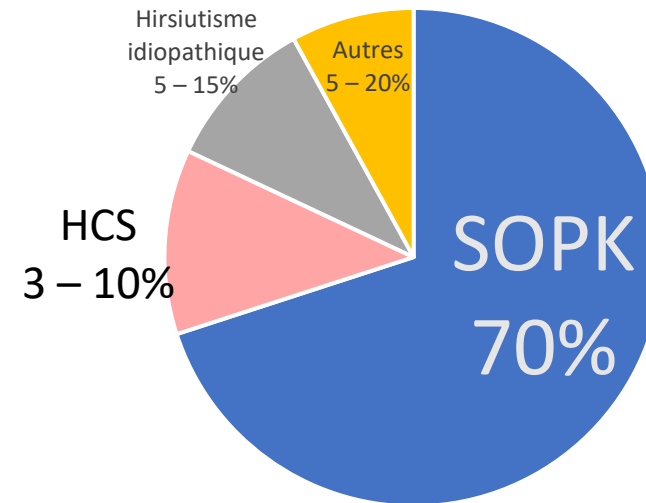
stic:
ou -

...de manière automatique, en apportant des exemples (couples X - Y)

Exemple 2 : Identification des hyperplasies congénitales des surrénales et diagnostic différentiel avec le SOPK

Contexte de l'hyperandrogénie et cas d'usage

- Excès d'androgènes circulants
- Origine ovarienne ou surrénalienne
- Prévalence : 8-12 %
- Tableau clinique hyperandrogénie
 - Hirsutisme
 - Acné
 - Troubles du cycle menstruel
 - Hypo/infertilité



Origine ovarienne Syndrome des ovaires polykystiques (SOPK)

- 1 femme sur 10
- Pathologie hétérogène, multifactorielle
- Une des plus grandes causes d'infertilité et d'irrégularités menstruelles

Origine surrénalienne Hyperplasie congénitale des surrénales (HCS-NC)

- Prévalence 1/200 à 1/1000
- **Bloc enzymatique (enzyme défaillante) voie biosynthèse des stéroïdes**
- 21-hydroxylase dans 95% cas

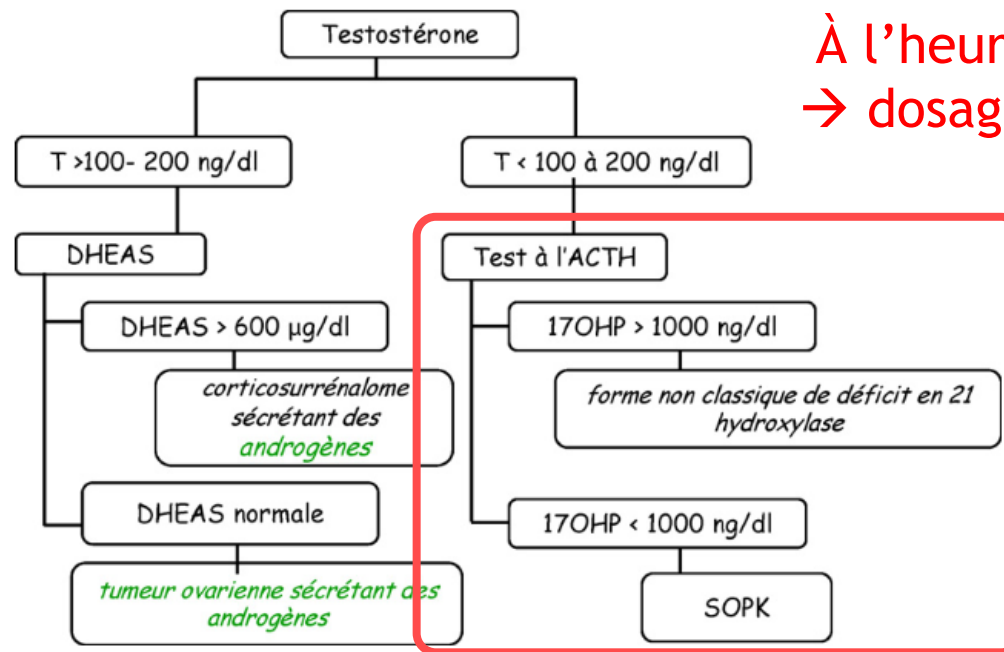
Exploration des hyperandrogénies



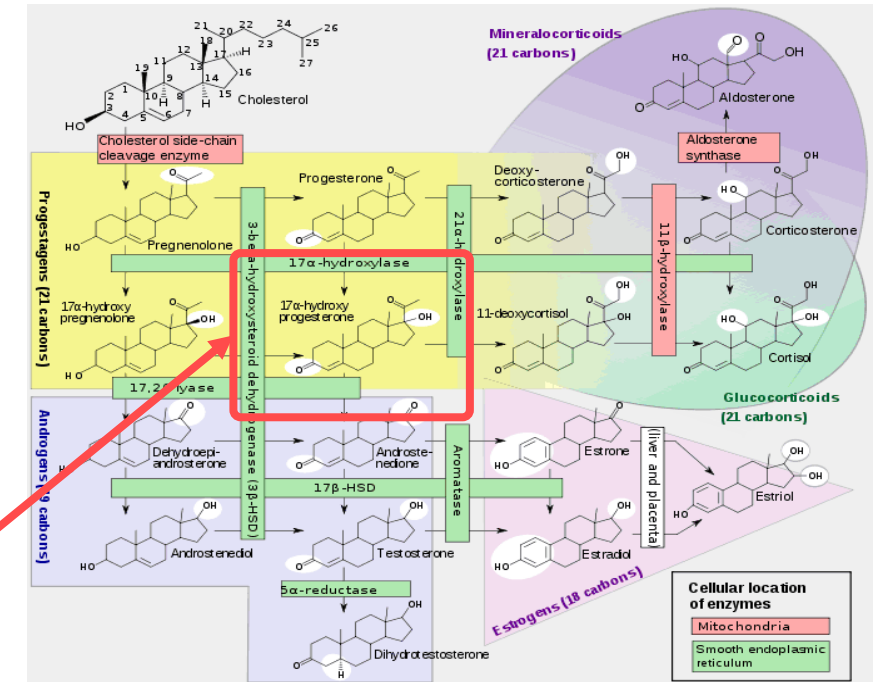
Annales d'Endocrinologie 71 (2010) e3–e8
Consensus de la Société française d'endocrinologie sur l'hyperandrogénie féminine

Recommandations pour l'exploration des hyperandrogénies

M. Pugeat ^{a,*}, H. Déchaud ^b, V. Raverot ^b, A. Denuzière ^b, R. Cohen ^c, P. Boudou ^d



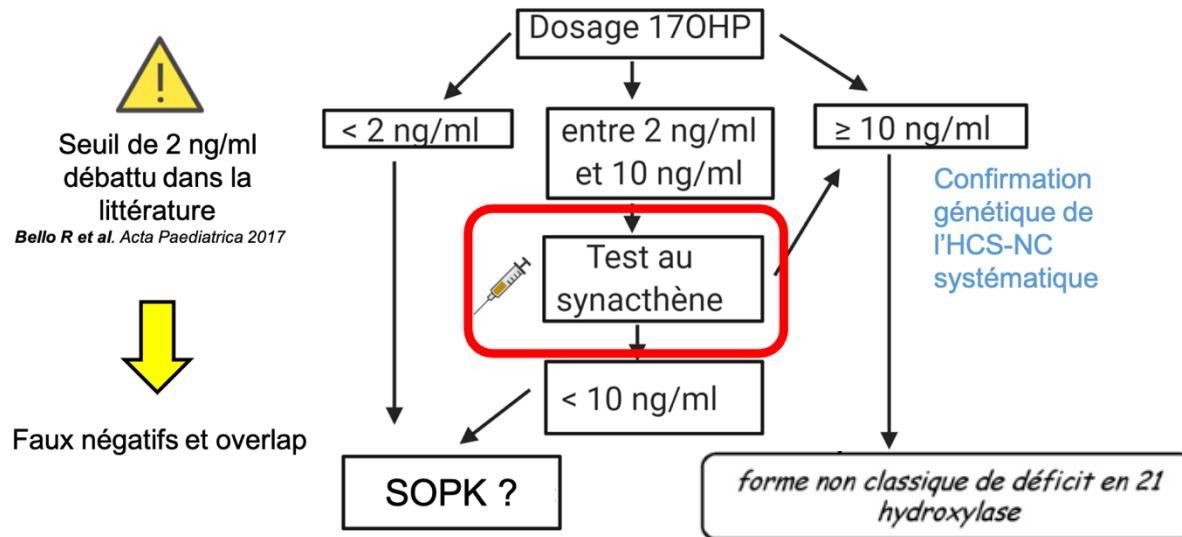
À l'heure actuelle
→ dosage [17-OHP]



Enjeu/problématique : distinguer correctement HCS-NC vs SOPK :

→ Présentation clinique identique

SOPK vs HCS-NC : les limites de l'épreuve dynamique



SOPK = 2 Critères de Rotterdam sur 3

International evidence-based guidelines for the assessment and management of polycystic ovary syndrome 2018

Irrégularité des cycles	> 35 jours ou < 21 jours
Hyperandrogénie clinique ou biologique	Acné, alopecie, hirsutisme Testostérone totale et SHBG
Aspect d'ovaires micropolykystiques à l'échographie	Volume ovarien > 10 mL ou > 20 follicules par ovaire



Mais...

→ Grande différence d'épidémiologies
SOPK >>> HCS-NC

→ Exploration dynamique en pratique



Elimination : differential diagnoses (TSH + prolactin + 17-OHP +/- others)

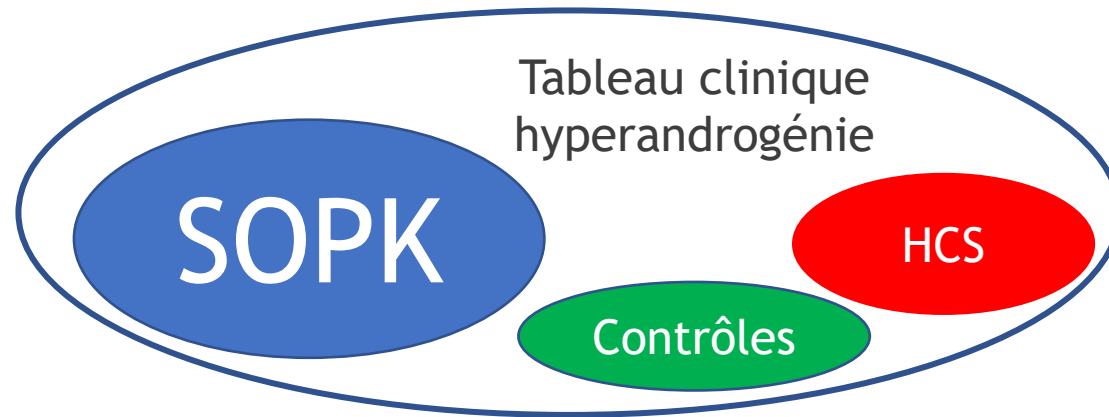
Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

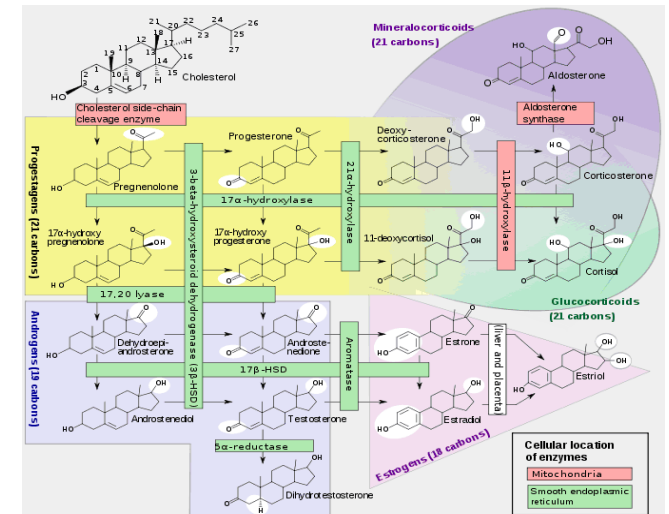
Question : peut-on se passer de l'épreuve dynamique ?



Profils stéroïdiens exhaustifs en taux de base (LC-MS/MS) = empreinte métabolique (15 stéroïdes)

Outils issus de Intelligence artificielle

Cartographie de la voie de biosynthèse des stéroïdes



Prélèvement unique

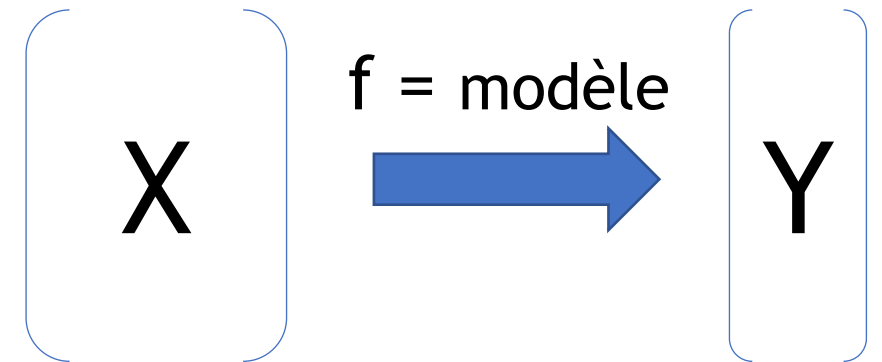


Cas d'usage de l'apprentissage automatique

→ Apprendre la fonction $f(X) = Y...$

→ Modélisation de la présence ou non d'une hyperplasie congénitale des surrénales (= phénomène) à partir d'une cartographie de la voie de biosynthèse des stéroïdes

→ OU prédire en avance le résultat de l'épreuve dynamique (pour l'éviter)



Caractéristiques d'un patient
= empreinte métabolique en LC-
MS/MS de 15 hormones

Etat/diagnostic:
= HCS-NC

...de manière automatique, en apportant des exemples (couples $X - Y$)

Introduction

Biopsie testiculaire

SOPK / CAH

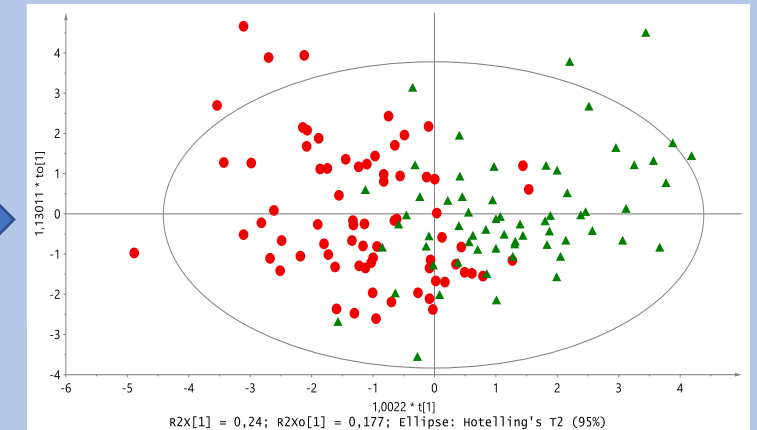
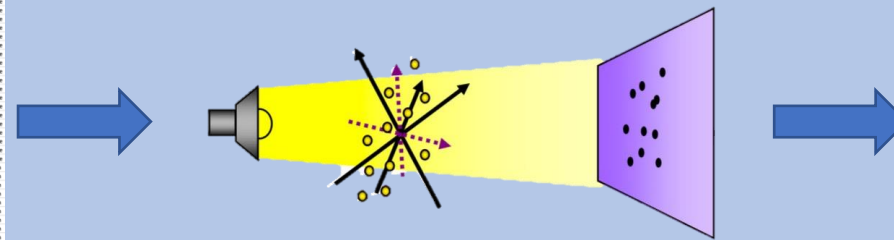
Conclusion

Algorithme OPLS-DA : comment ça marche ?

Modèle = OPLS-DA (*Orthogonal Partial Least Square Discriminant Analysis*)

- ✓ Données manquantes
- ✓ Colinéarités
- ✓ Output très visuels
- ✓ Bonne intelligibilité / interprétabilité du modèle

Patient	Abdominale	Costale	Libre	21DP	Costoverbra	Dorsal	DOE	Testostérone	17OHtestostérone	DHEA	17OHtestostérone	Progesterone	Androstène	Classe	
Widmer_Patient	0.082	25.435	113.371	1.054	0.34	1.095	2.343	0.043	0.307	9.817	4.206	0.17	8.791	2.294	Patiente
Widmer_Patient	0.092	23.084	103.728	2.099	0.485	0.861	2.477	0.025	0.387	10.115	4.125	0.17	0.263	1.226	Patiente
Dagga_Patient	0.16	27.293	110.176	1.521	0.334	1.427	3.701	0.047	0.092	5.525	11.338	0.381	0.288	3.384	Patiente
Karlskron_Patient	0.095	25.12	105.047	2.044	1.138	1.849	1.871	0.05	0.709	6.382	12.841	1.491	0.984	5.025	Patiente
Karlskron_Patient	0.069	24.853	97.933	1.405	2.219	0.957	2.907	0.032	0.452	13.687	23.31	1.56	0.346	5.003	Patiente
Kozsdy_Patient	0.133	29.222	147.644	7.64	1.751	0.788	2.615	0.009	0.387	5.463	12.23	0.368	0.166	1.217	Patiente
Laraki_Patient	0.092	27.563	117.353	2.367	0.933	0.973	3.539	0.01	0.691	4.509	12.758	0.883	0.174	2.04	Patiente
Laraki_Patient	0.144	33.19	96.353	2.183	4.552	0.991	0.884	0.008	0.302	12.199	1.376	0.385	0.473	1.798	Patiente
Laraki_Patient	0.208	21.943	113.242	1.36	0.247	1.852	2.306	0.068	0.315	4.263	6.389	2.301	16.44	4.031	Patiente
Desmet_Patient	0.013	22.16	128.51	1.799	1.792	1.154	2.185	0.086	0.149	34.27	2.104	0.288	1.147	4.325	Patiente
Desmet_Patient	0.142	17.467	82.883	1.113	1.919	0.81	1.007	0.017	0.632	9.582	10.115	5.934	0.253	3.842	Patiente
Desmet_Patient	0.023	18.395	106.312	3.028	2.356	1.145	1.293	0.017	0.242	10.098	7.75	0.69	0.248	1.146	Patiente
Marx_Patient	0.051	18.779	78.432	2.261	0.493	0.61	2.822	0.064	0.188	3.593	11.339	0.706	0.221	1.029	Patiente
Marx_Patient	0.029	27.754	112.682	1.318	0.178	1.184	1.208	0.004	0.185	1.111	6.421	0.329	0.071	0.767	Patiente
Luffmann_Patient	0.06	29.89	92.69	3.05	0.787	0.57	1.44	0.01	0.44	9.51	5.49	2.75	0.25	1.05	Patiente
Luffmann_Patient	0.01	28.41	119.75	3.47	3.436	0.54	2.49	0.01	0.44	8.2	14.59	3.1	0.2	4.97	Patiente
Yoon_Patient	0.22	15.78	68.83	4.121	3.488	2.22	1.79	0.06	1.28	38.49	11	3.18	1.24	1.23	Patiente
Yoon_Patient	0.05	22.76	129.58	2.44	1.512	1.87	1.21	0.04	0.27	4.49	4.71	0.14	1.87	2.99	Patiente
Yoon_Patient	0.12	12.56	94.15	10.51	10.119	2.35	8.39	0.06	0.09	33.75	26.84	14.51	1.11	10.57	Patiente
Hahn_Patient	0.094	19.759	60.471	0.512	0.006	1.255	1.099	0.015	0.221	0.478	4.369	0.001	0.109	2.134	Témoin
Hahn_Patient	0.134	17.884	69.354	0.986	0.008	1.59	1.512	0.022	0.292	0.547	4.405	0.451	0.101	1.313	Témoin
Dagga_Patient	0.099	20.001	100.255	0.907	0.008	4.799	0.989	0.043	0.293	0.811	4.186	0.135	7.632	1.485	Témoin
Dagga_Patient	0.018	25.091	129.871	0.921	0.019	3.268	1.199	0.01	0.266	0.975	10.777	2.127	0.074	3.18	Témoin
Chen_Patient	0.041	21.772	91.007	0.647	0.006	1.809	2.198	0.044	0.181	0.949	5.242	0.963	0.122	1.05	Témoin
Chen_Patient	0.141	25.601	101.144	0.989	0.027	3.459	0.689	0.034	0.194	0.999	6.882	0.144	1.127	1.245	Témoin
Lang_Patient	0.038	25.127	86.135	2.01	0.021	2.511	1.2	0.031	0.187	0.467	4.884	0.178	0.033	1.053	Témoin
Lang_Patient	0.137	24.448	119.91	1.302	0.024	0.55	1.72	0.01	0.226	0.713	8.49	0.738	0.884	1.86	Témoin
Günz_Patient	0.08	21.144	179.211	1.428	0.079	16.57	0.472	0.007	0.146	0.108	2.414	0.139	0.055	1.087	Témoin
Widmer_Patient	0.059	21.345	122.643	1.897	0.027	5.224	0.879	0.047	0.14	1.363	5.443	2.238	5.718	1.822	Témoin
Widmer_Patient	0.044	21.257	102.744	0.959	0.019	0.478	0.927	0.009	0.099	0.297	1.162	0.26	1.124	1.124	Témoin



Données brutes

Projection

Score plot

Introduction

Biopsie testiculaire

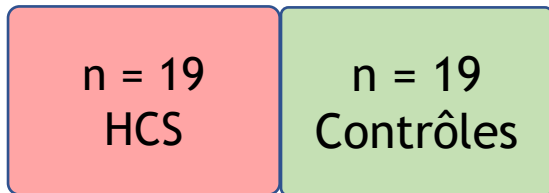
SOPK / CAH

Conclusion

En pratique : 2 Cohortes - étude bicentrique

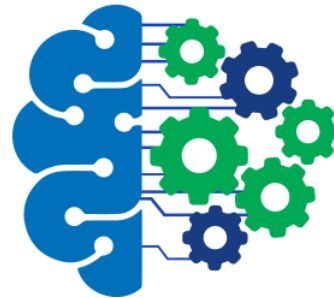


Pr. Jacques Young



Cohorte d'entraînement

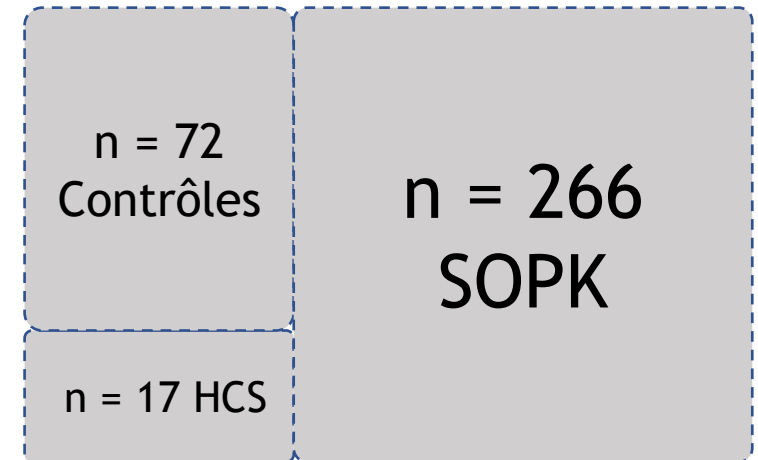
Construction du modèle



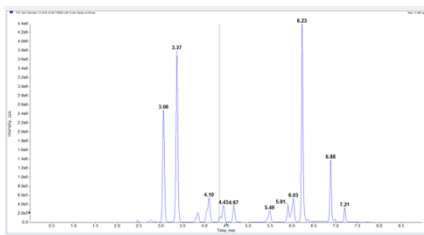
Entraîné à reconnaître
des empreintes
métaboliques de HCS



Pr. Anne Bachelot



Cohorte de test externe
= évaluer ses performances,
sa capacité à exclure les HCS



Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

En pratique : 2 Cohortes - étude bicentrique

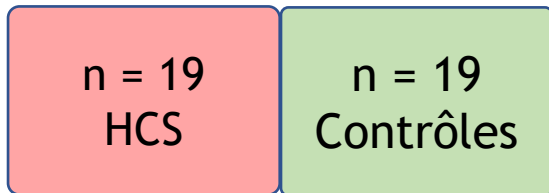


Pr. Jacques Young

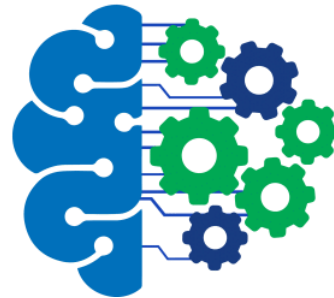


Pr. Anne Bachelot

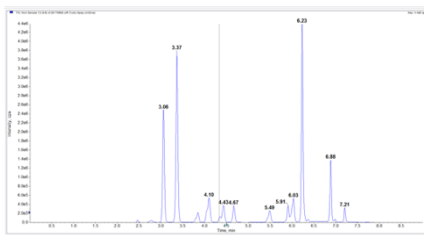
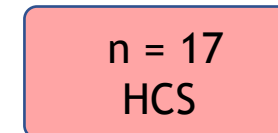
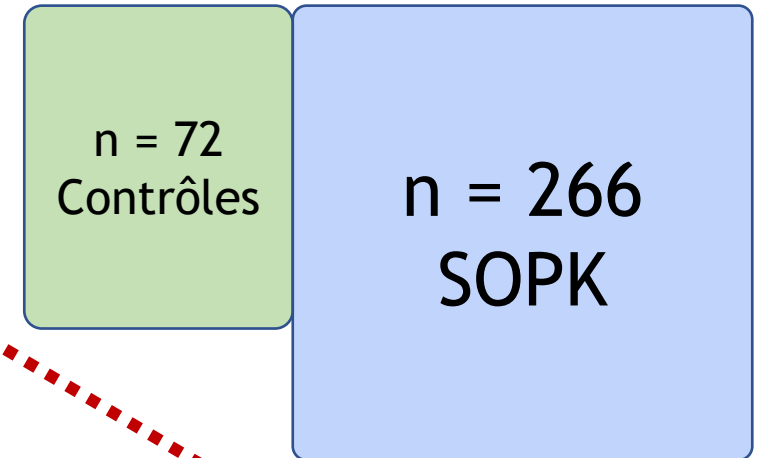
Construction du modèle



Cohorte d'entraînement



Entraîné à reconnaître
des empreintes
métaboliques de HCS



Introduction

Biopsie testiculaire

SOPK / CAH

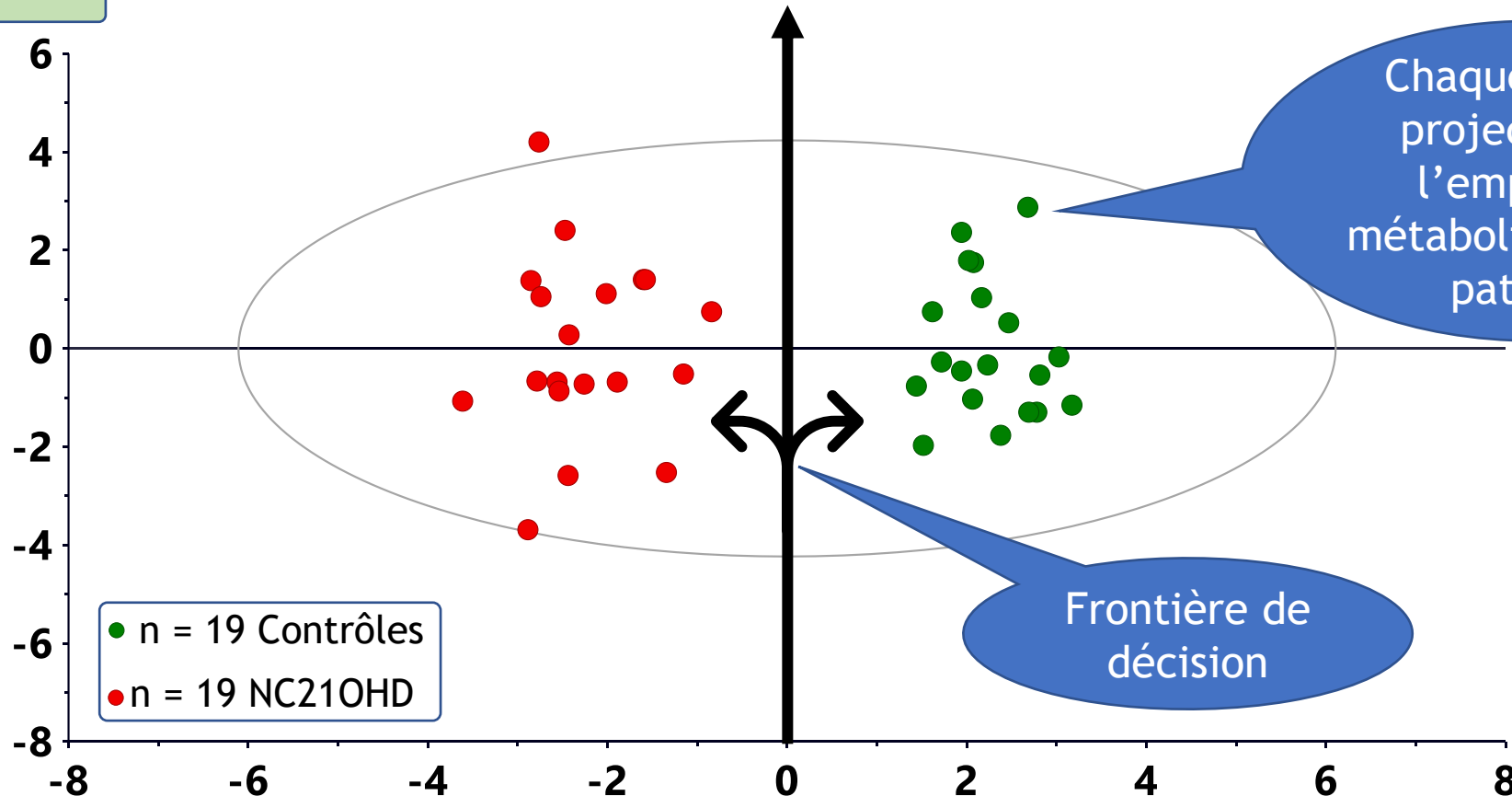
Conclusion

Construction du modèle avec la cohorte d'entraînement

n = 19
HCS

n = 19
Contrôles

Ce plan = représentation graphique du modèle



Introduction

Biopsie testiculaire

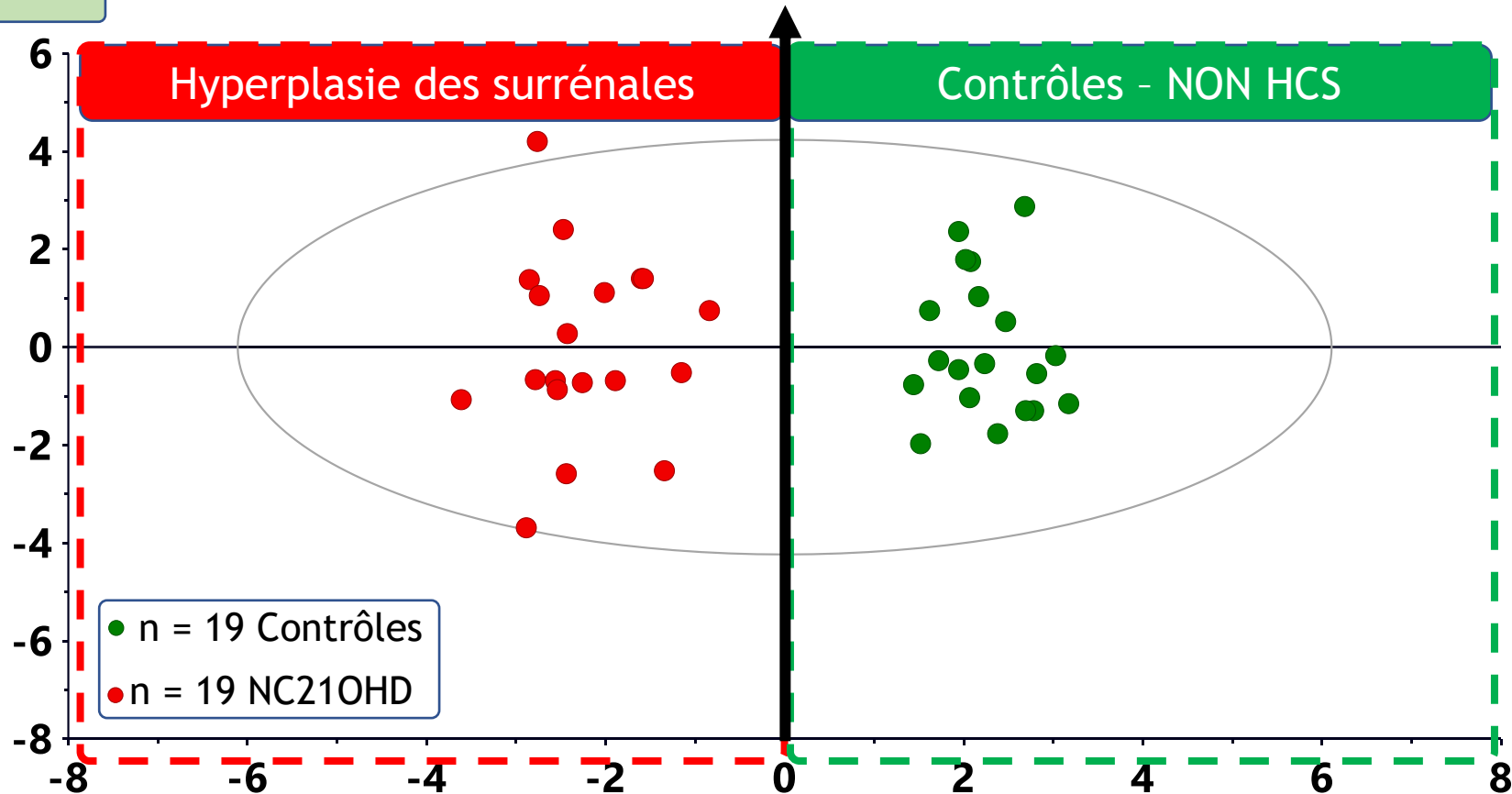
SOPK / CAH

Conclusion

Dessin de 2 « zones » : HCS et contrôle

n = 19
HCS

n = 19
Contrôles



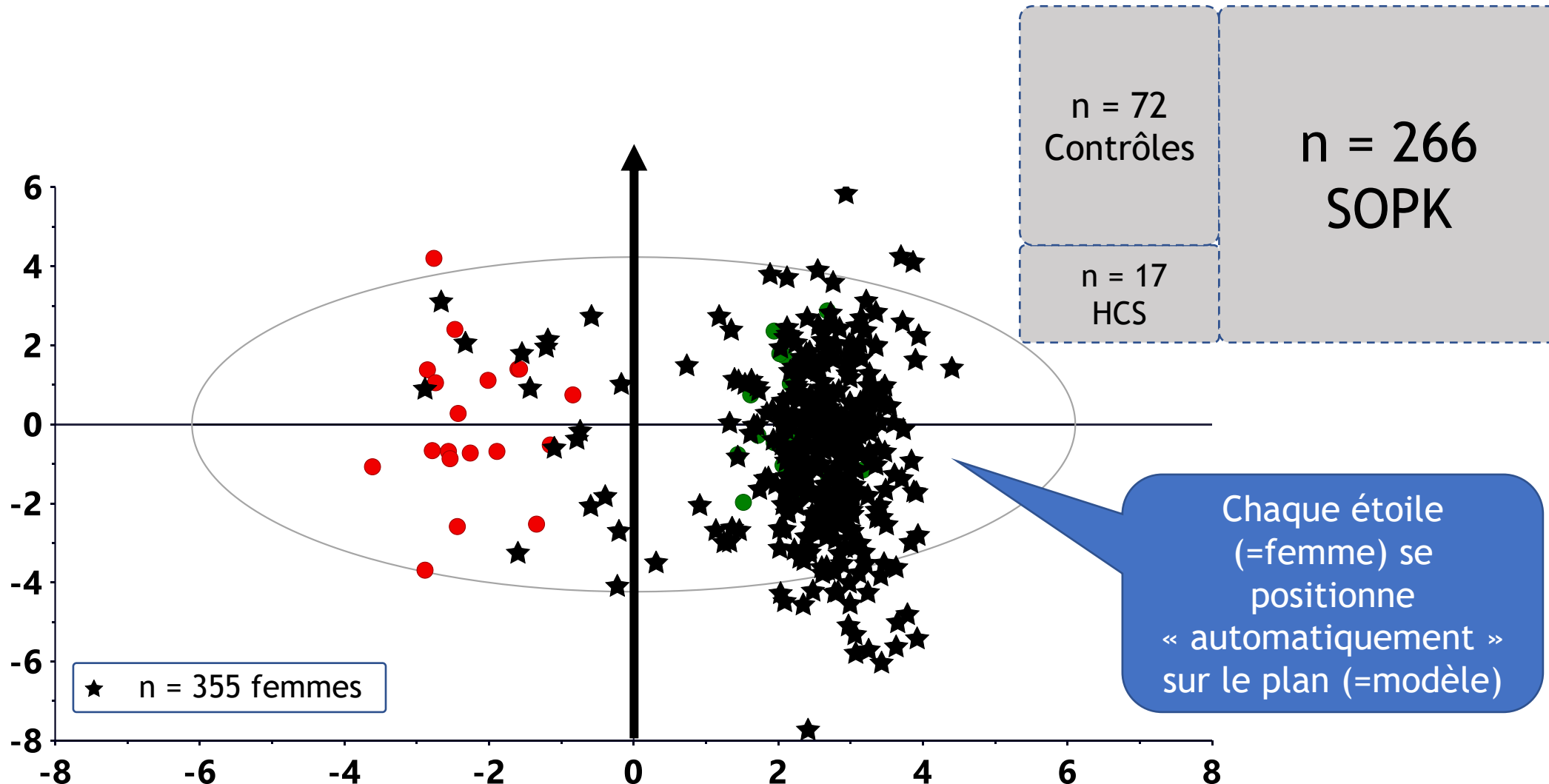
Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

Injection des empreintes/patientes de la cohorte de test dans le modèle « à l'aveugle »



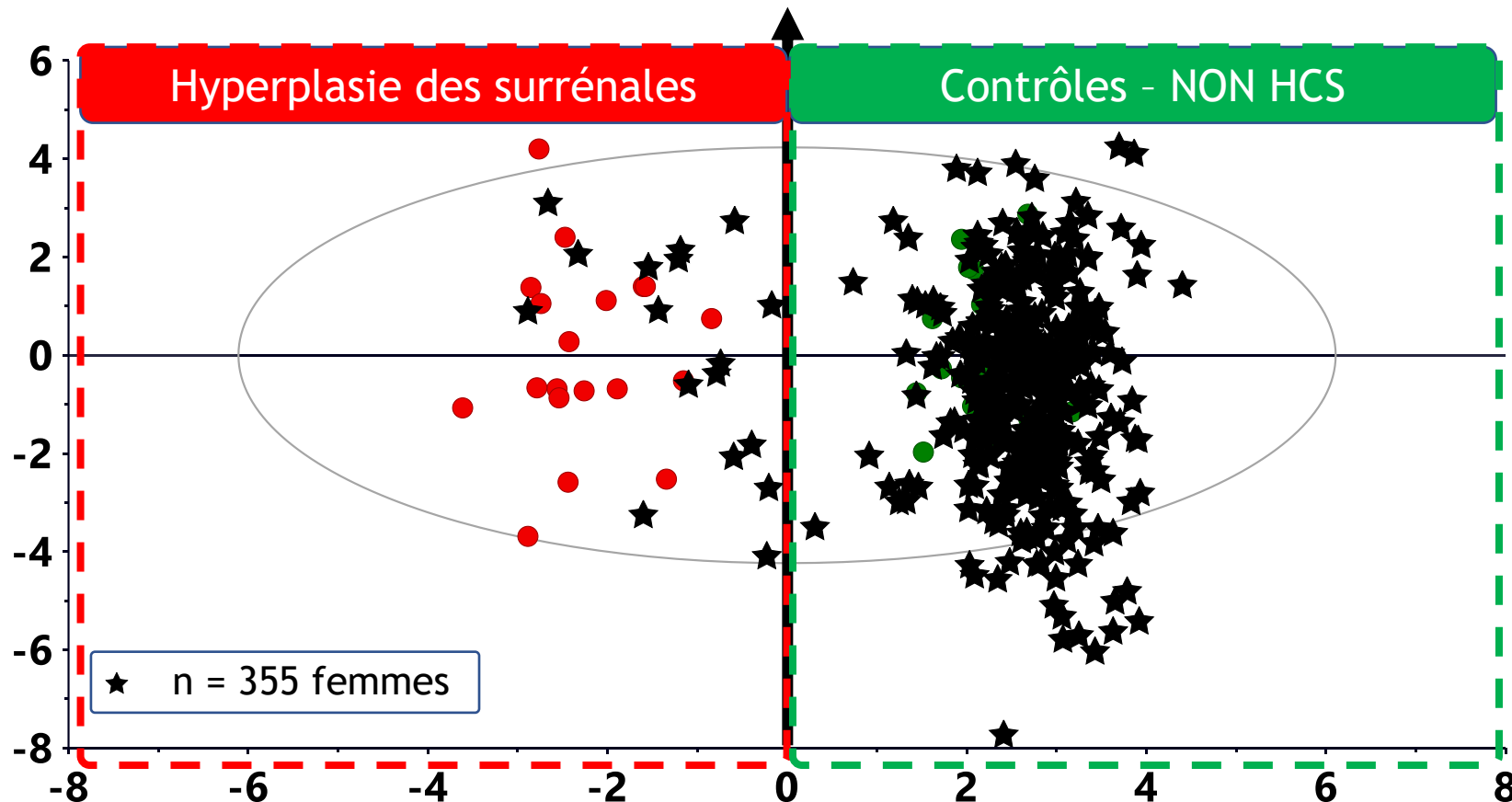
Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

Comparaison position de l'étoile (score donné par le modèle) et la couleur de l'étoile (statut réel)



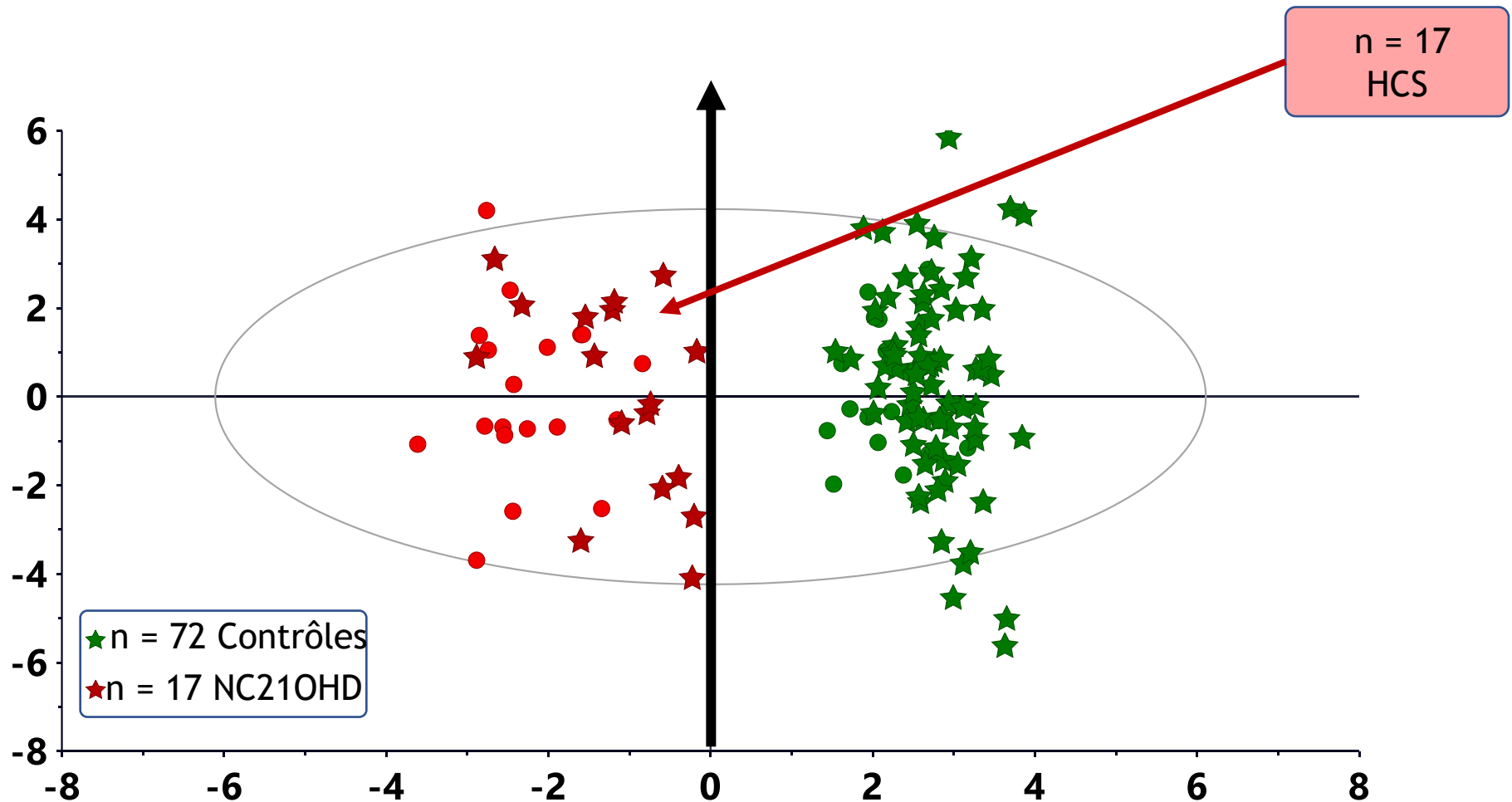
Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

Patientes avec déficit en 21-hydroxylase sont toutes bien positionnées et bien classées par le modèle



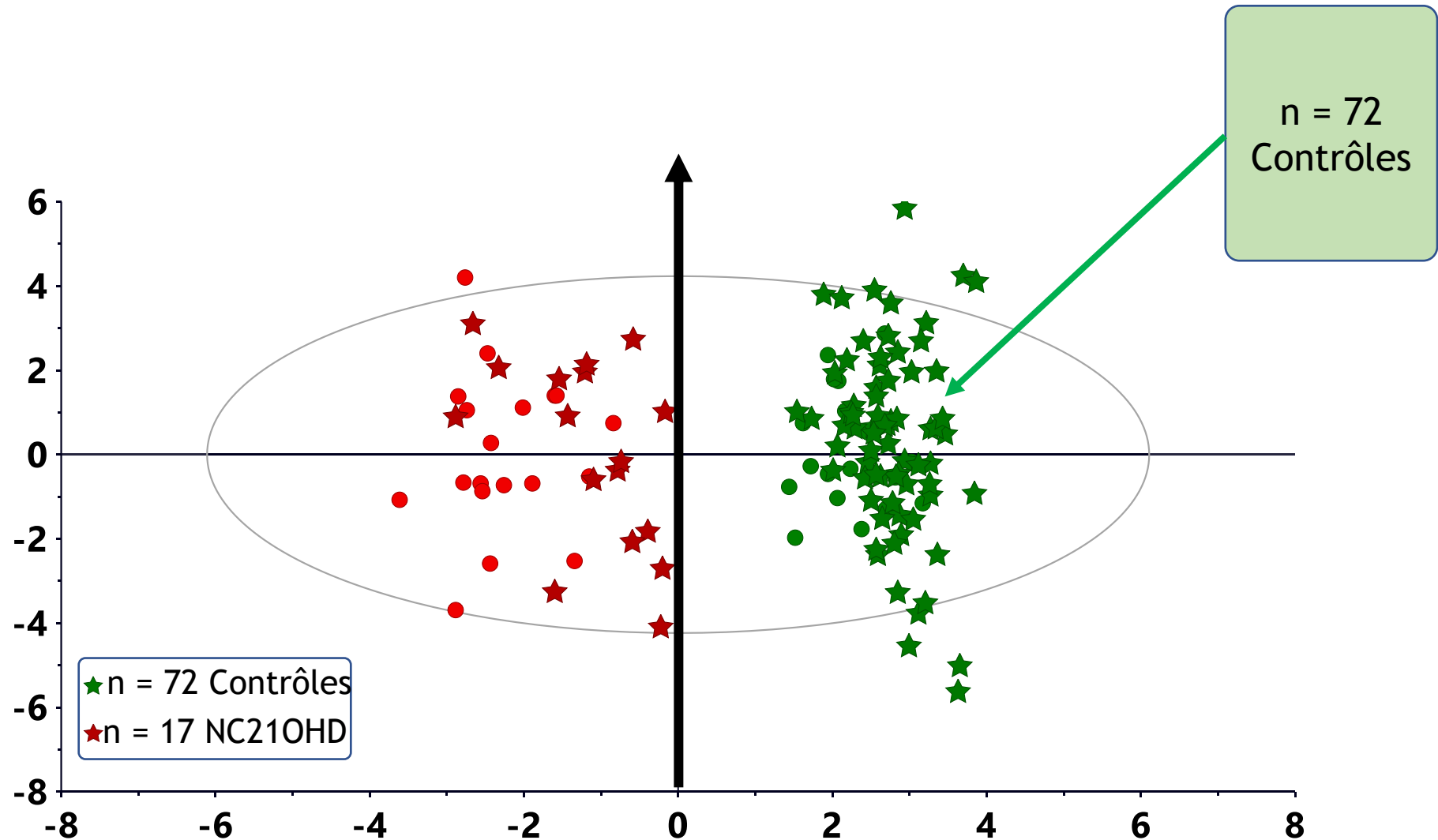
Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

Les contrôles sont toutes bien positionnées dans la zone « non-HCS »



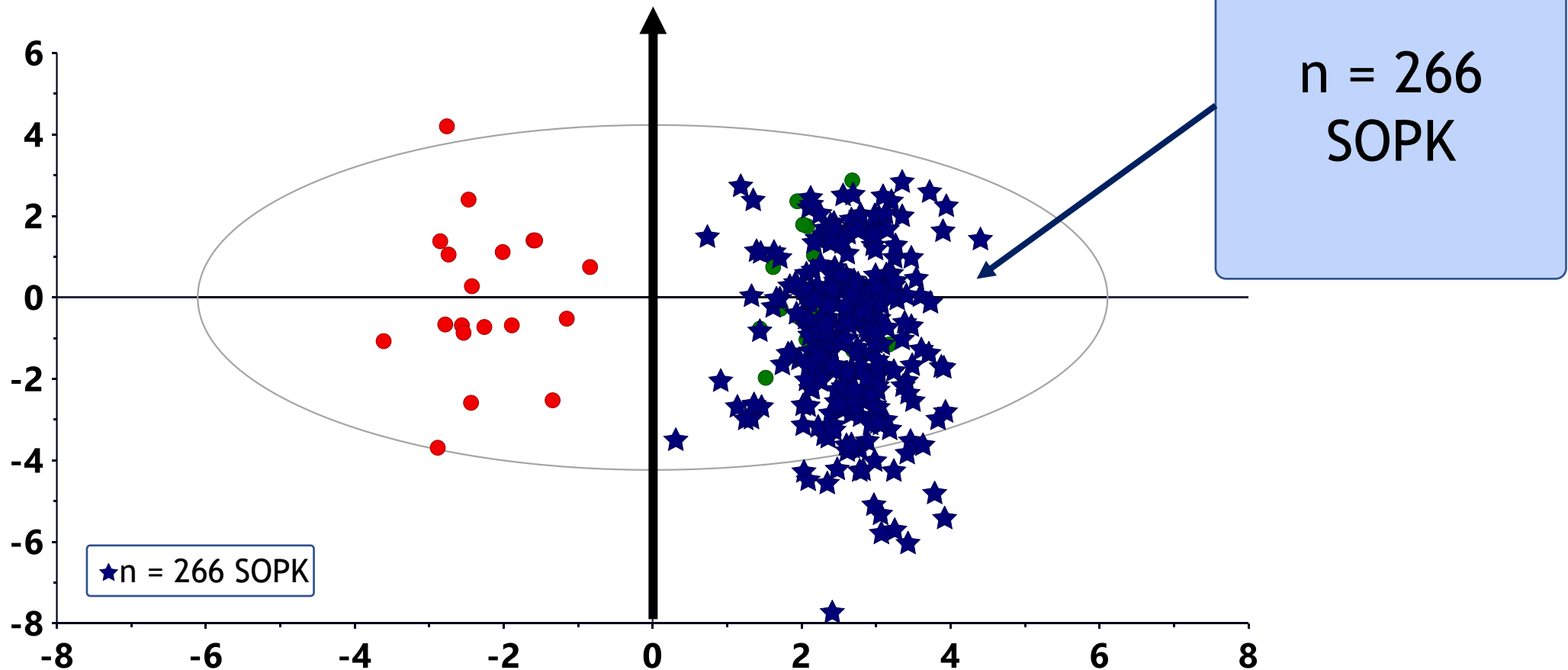
Introduction

Biopsie testiculaire

SOPK / CAH

Conclusion

Patientes SOPK sont placées du côté « contrôle »
→ bonne ségrégation entre HCS et SOPK malgré clinique similaire !



Introduction

Biopsie testiculaire

SOPK / CAH

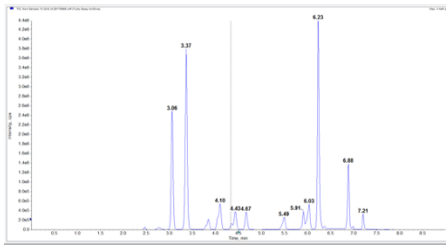
Conclusion



n = 72
Contrôles

n = 17
HCS

n = 266
SOPK



Diagnostic

n = 72
Contrôles

n = 266
SOPK

n = 17
HCS

Performances du modèle Stratification de toutes les patientes NC21OHD

Sensibilité et spécificité 100%

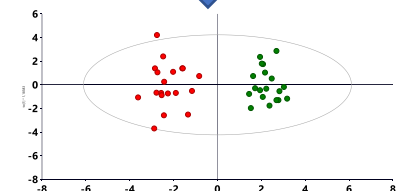
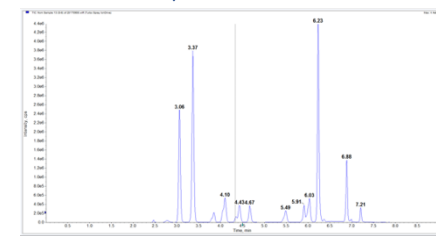
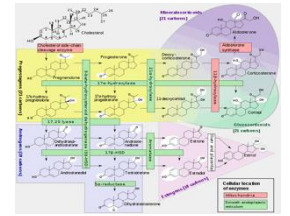
Modèle OPLS-DA		Diagnostic		
		non-NC21OHD	NC21OHD	
Prédiction	non-HCS	338	0	338
	HCS	0	17	17
		338	17	355

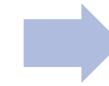
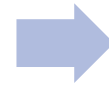
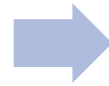
Résultat indépendant du timing du prélèvement au cours du cycle
n = 32 patientes prélevées en phase lutéale ([P] > 10 ng/mL)

La structure de la voie métabolique suffisamment informative
Pour « absorber » les effets du cycle
➔ Confirme l'hypothèse de départ

Conclusion exemple 2

- **Oui**, il est possible de s'affranchir de l'épreuve dynamique pour exclure le diagnostic d'HCS dans un contexte d'hyperandrogénie
- Meilleures performances que l'exemple 1 + méthodologie plus robuste (bicentrique)
- Pourquoi ?
 - Matrice X = la voie métabolique directement impliquée et responsable de la maladie (monogénique due à bloc enzymatique)
 - Proximité ++++ entre phénomène à modéliser et les données recueillies : on est « au cœur du problème »
 - Label Y : génétique (mutations bien identifiées)





Conclusion pour la construction d'un modèle d'IA

1. Cas d'usage : véritable problématique ?

- Automatiser une tâche répétitive, pénible...
- Répondre à une question sans réponse

2. Partir d'hypothèses solides pour constituer la « matrice X »

- Obtenir des données suffisamment explicatives/informatives (ex : longueur des pieds pour prédire K poumon) - modèles simples suffisants et moins d'exemples nécessaire !

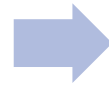
3. Pouvoir étiqueter (Y) de manière fiable (sinon le modèle apprend n'importe quoi !)

- Seul un expert du domaine d'étude peut le faire !

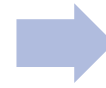
4. Une fois qu'on a défini le X et le Y : quels exemples prendre ? Exemples représentatifs ?

- Design de l'étude +++ attention aux données rétrospectives !

Introduction



Biopsie testiculaire



SOPK / CAH



Conclusion

Merci pour votre attention