

Aplicación del algoritmo SCAN en el agrupamiento de imágenes de trayectorias espermáticas: Identificación de la heterogeneidad de la respuesta espermática a la Ketanserina

Eder Alejandro Rodríguez Martínez¹, Cindy Ursula Rivas Arzaluz²,
Andrés Aragón Martínez²

¹ University of Picardie Jules Verne,
France

² Universidad Nacional Autónoma de México,
Facultad de Estudios Superiores Iztacala,
México

armandres@gmail.com

Resumen. La identificación de subpoblaciones cinemáticas es de vital importancia para comprender la naturaleza biológica de la heterogeneidad de los espermatozoides. Actualmente, los datos de parámetros de movilidad obtenidos mediante un sistema de análisis espermático asistido por computadora (CASA) se han utilizado como entrada para diferentes algoritmos con el fin de identificar subpoblaciones cinemáticas. En contraste, las imágenes de las trayectorias solamente se han representado como ejemplos de los patrones de movilidad en cada subpoblación. En este estudio, se escribió un código en Python para reconstruir las imágenes de las trayectorias a partir de sus coordenadas; estas se utilizaron como entrada para el algoritmo de SCAN; y, se describió estadísticamente el efecto de la ketanserina en la cinemática de cada grupo identificado. Las imágenes de las trayectorias en cada grupo se mostraron de una manera que denominamos gráfico de Pollock. Se trataron muestras de semen de cerdo con distintas concentraciones de ketanserina y se utilizaron muestras no tratadas como control. La movilidad espermática en cada muestra se analizó a los 0 y 30 minutos de incubación. Se obtuvieron seis grupos (subpoblaciones). La subpoblación 2 presentó los valores más altos de velocidad espermática a los 0 y 30 minutos. Después de 30 minutos de incubación, la ketanserina indujo un aumento de los valores de la velocidad curvilínea a altas concentraciones, mientras que la linealidad y la velocidad rectilínea disminuyeron. Nuestro flujo de trabajo permite una mejor identificación de las subpoblaciones cinemáticas que el enfoque tradicional y proporciona información sobre la heterogeneidad de la respuesta a la ketanserina. Por lo tanto, este podría impactar significativamente en la investigación de la relación entre la heterogeneidad de los espermatozoides y la fertilidad. Adicionalmente, el código se puede encontrar en¹.

Palabras clave: Subpoblaciones cinemáticas, imágenes de trayectorias espermáticas, análisis espermático asistido por computadora, patrones de movilidad.

Application of the SCAN Algorithm in the Clustering of Sperm Trajectory Images: Identification of the Heterogeneity of the Sperm Response to Ketanserin

Abstract. Identifying kinematic subpopulations is of vital importance for understanding the biological nature of sperm heterogeneity. Currently, mobility parameter data obtained through a computer-assisted sperm analysis system (CASA) has been used as input for different algorithms to identify kinematic subpopulations. In contrast, trajectory images have only been represented as examples of mobility patterns in each subpopulation. In this study, Python code was written to reconstruct trajectory images from their coordinates; then, the trajectory images were used as input for the SCAN algorithm; finally, the effect of ketanserin on the kinematics of each identified group was statistically described. The trajectory images in each group were displayed in a way we called Pollock plots. Pig semen samples were treated with different concentrations of ketanserin and untreated samples were used as controls. Sperm motility in each sample was analyzed at 0 and 30 minutes of incubation. Six clusters (subpopulations) were obtained. Subpopulation 2 had the highest sperm velocity values at 0 or 30 minutes. After 30 minutes of incubation, ketanserin induced an increase in curvilinear velocity values at high concentrations, while linearity and straight-line velocity decreased. Our workflow allows for better identification of kinematic subpopulations than the traditional approach and provides information on the heterogeneity of the response to ketanserin. Therefore, this could significantly impact research on the relationship between sperm heterogeneity and fertility. Additionally, the code can be found at¹.

Keywords: Sperm subpopulations, images of sperm trajectories, computer assisted sperm analysis, patterns of motility.

1. Introducción

Actualmente, el procedimiento estadístico para identificar subpoblaciones cinemáticas utiliza los datos de parámetros de movilidad, obtenidos de un sistema de análisis de espermatozoides asistido por computadora (CASA), para identificar grupos de datos con base en su similitud [1, 2].

Una vez que se identifican y describen estadísticamente las subpoblaciones, generalmente se presentan imágenes representativas de las trayectorias seguidas por algunos espermatozoides, en cada subpoblación [3, 4]. Sin embargo, estos enfoques no utilizan toda la información potencial proporcionada por los sistemas CASA, es decir, los valores de los parámetros de movilidad y las trayectorias seguidas por los espermatozoides.

El software de los sistemas CASA rastrea cada espermatozoide siguiendo el centroide de los píxeles que definen la cabeza; posteriormente, un algoritmo calcula las velocidades y otros parámetros de movilidad de cada espermatozoide, basándose en las coordenadas del centroide.

¹ github.com/eder1234/traj_img_clusters

A continuación, diferentes técnicas estadísticas utilizan estos parámetros de movilidad para identificar subpoblaciones cinemáticas [5]. Aunque estos parámetros proporcionan información relevante relacionada con la movilidad de los espermatozoides, no son suficientes para, de manera inversa, reconstruir la trayectoria representada por las coordenadas a lo largo del tiempo.

En este sentido, podemos afirmar, entonces, que los parámetros de movilidad están asociados con la trayectoria del espermatozoide y que la trayectoria puede reconstruirse a partir de las coordenadas. En principio, las imágenes de las trayectorias espermáticas (también llamadas patrones de movilidad) [5] son en realidad una colección heterogénea de trayectorias, como se puede apreciar en las imágenes de trayectorias y en los valores de los parámetros de movilidad asociados a ellas [5, 6].

Hasta ahora, las imágenes de las trayectorias no se habían considerado como una entrada para algoritmos de aprendizaje de máquinas, que permiten identificar subpoblaciones espermáticas. Esta situación abre una ventana de oportunidad para desarrollar nuevos enfoques que permitan comprender, cuantitativamente, los patrones de movilidad espermática.

En consecuencia, la información biológica subyacente en los datos obtenidos de los sistemas CASA podría explotarse aún más. En primer lugar, describiendo finamente la cinemática de las subpoblaciones y, en segundo lugar, mejorando la interpretación de la movilidad de los espermatozoides en condiciones fisiológicas y experimentales.

La identificación de subpoblaciones cinemáticas proporciona información valiosa sobre la fisiología de los espermatozoides. Sin embargo, en la mayoría de los estudios en los que se han utilizado sistemas CASA no se han identificado subpoblaciones cinemáticas. Por otra parte, se ha reportado que los espermatozoides expresan proteínas relacionadas con la comunicación serotoninérgica, como receptores, transportadores y proteínas metabolizadoras [7]. En ese artículo, los autores describieron un aumento en las velocidades promedio de los espermatozoides debido a la exposición a la serotonina.

En otros estudios se describieron los valores promedio de los parámetros de movilidad de los espermatozoides expuestos a agonistas o antagonistas de los receptores de serotonina [8, 9]. Actualmente, se desconoce la estructura de las subpoblaciones cinemáticas de espermatozoides expuestos a sustancias que regulan la comunicación serotoninérgica. Sin embargo, si la serotonina estimula la movilidad de los espermatozoides, entonces el bloqueo de la comunicación serotoninérgica podría cambiar la estructura de las subpoblaciones cinemáticas.

En un acercamiento más reciente, el problema se abordó por medio de un agrupamiento de imágenes de trayectorias ya que estas últimas contienen información relevante que no es completamente descrita por los parámetros de movilidad [10]. Específicamente, la tarea de agrupamiento la llevó a cabo el algoritmo aglomerativo jerárquico, y el efecto de la ketanserina fue analizado estadísticamente en cada subpoblación espermática.

Con respecto a otros métodos de agrupamiento llevado a cabo por expertos [10], el aglomerativo jerárquico es un algoritmo objetivo y automático que no requiere de intervención humana. Sin embargo, una objeción que puede hacerse al uso del algoritmo aglomerativo jerárquico es que no es robusto a la rotación.

Por lo tanto, en este trabajo se optó por la implementación del algoritmo de Agrupamiento Semántico por Adaptación de los Vecinos más Próximos (o SCAN, por sus siglas en inglés) [11] para agrupar las imágenes de las trayectorias espermáticas. El

algoritmo SCAN es robusto a transformaciones de imágenes deseadas, por ejemplo, la rotación, la escala y la translación, entre otras.

Esto se debe a que la función de costo es capaz de destruir la información no deseada relativa a las transformaciones de imagen. En consecuencia, los agrupamientos resultantes son más homogéneos que los propuestos por el algoritmo aglomerativo jerárquico.

El objetivo de este trabajo fue desarrollar un método computacional, basado en algoritmos de aprendizaje profundo, para identificar subpoblaciones basadas en imágenes de trayectorias espermáticas. Este método computacional permite identificar los cambios estructurales en las subpoblaciones cinemáticas de espermatozoides de cerdo expuestas a ketanserina.

2. Materiales y métodos

Se utilizaron diez dosis de semen, dos dosis de cada uno de cinco cerdos fértiles; las dosis se obtuvieron del Centro de Enseñanza, Investigación y Extensión en Producción Porcina de la Facultad de Medicina Veterinaria y Zootecnia (UNAM). Las dosis de semen se produjeron inicialmente para inseminación artificial, fueron transportadas y procesadas de acuerdo con [12].

La unidad experimental en este trabajo fue la alícuota. De cada dosis de semen, se tomaron cuatro alícuotas ($N = 40$ alícuotas); 10 de estas dosis no se trataron y sirvieron como control, y el resto se trató con distintas concentraciones de ketanserina, según se especifica a continuación.

Se utilizó un sistema CASA de código abierto. La movilidad fue evaluada siguiendo el método descrito en [12]. De manera similar a [13], se evaluó visualmente la trayectoria de cada espermatozoide identificado y registrado en cada campo antes del análisis de la secuencia de seguimiento. Esta consideración se llevó a cabo para eliminar posibles restos celulares y disminuir el riesgo de incluir trayectorias poco claras en el análisis.

Se registraron los valores cinemáticos de cada uno de los siguientes parámetros de movilidad para cada espermatozoide analizado: velocidad promedio del recorrido (VAP, $\mu\text{m/s}$), velocidad curvilínea (VCL, $\mu\text{m/s}$), velocidad en línea recta (VSL, $\mu\text{m/s}$), frecuencia de cruce de batido (BCF, Hz), linealidad (LIN, VSL/VCL), coeficiente de rectitud (STR, VSL/VAP), amplitud de desplazamiento lateral de la cabeza (ALH, μm) y bamboleo (WOB, VAP/VCL).

El plugin CASA_bgm genera una hoja de datos con los valores cinemáticos de cada espermatozoide y muestra una imagen donde se describen las trayectorias de todos los espermatozoides analizados en una secuencia de video [13].

El plugin permite guardar las coordenadas de cada espermatozoide analizado. Para automatizar la construcción de los dataset para las coordenadas de los espermatozoides y los datos cinemáticos individuales, se escribió un script en Python que toma los datos de la hoja de resultados como entrada y luego construye dos archivos csv: uno para las coordenadas y otro para los parámetros de movilidad (el script en Python se nombró `log_motility_data-6.py` y está disponible en doi: 10.7910/DVN/YI3N4Q). El archivo de coordenadas se utilizó como entrada para otros scripts en Python, que convirtieron las coordenadas en imágenes individuales.

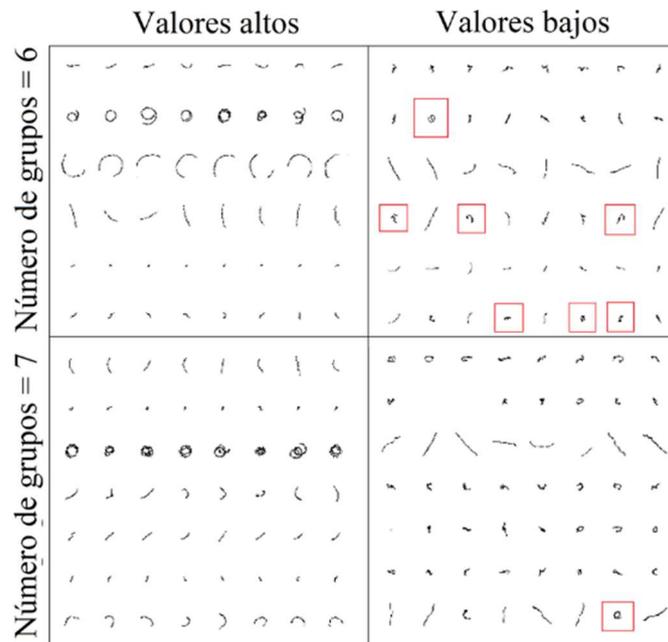


Fig. 1. Eficacia del agrupamiento de imágenes de trayectorias realizadas por el algoritmo SCAN. En vertical, se muestran los resultados del agrupamiento realizado por SCAN, con un ajuste a 6 y 7 grupos. En horizontal, se muestran las imágenes más cercanas (Valores altos) y más alejadas (Valores bajos) del centro de cada grupo al que pertenecen. Los cuadrados rojos señalan a las trayectorias semánticamente mal agrupadas.

Las imágenes de trayectorias se generaron como imágenes png utilizando matplotlib versión 3.6. Se utilizó el módulo de imagen Pillow Imaging Library para convertir cada trayectoria en un vector binario; este módulo tiene la opción "L" con la que los píxeles se asignan a su límite más cercano, ya sea negro (0) o blanco (256) [14].

Se verificó que los datos fueran binarios y se establecieron todos los valores como 1; por lo tanto, nuestros datos de entrada fueron un conjunto de vectores binarios con valores de 0 o 1. A partir de ahí, las imágenes individuales se almacenaron en una carpeta (el script se llamó traj-ah-6-full-0.ipynb y está disponible como un Jupyter notebook en doi:10.7910/DVN/CBMKVA).

Una vez que las imágenes de trayectorias fueron generadas y almacenadas, el siguiente paso consistió en clasificarlas o agruparlas automáticamente. En resumen, el problema consiste en agrupar las imágenes de forma que cada imagen sea semánticamente consistente con las otras que pertenecen al mismo grupo.

En la literatura, existen diferentes algoritmos de agrupamiento de imágenes, ej. Agrupamiento Selectivo Pseudo-etiquetado (o SPC, por sus siglas en inglés) [15], Agrupamiento Pseudo-etiquetado de Imágenes (o SPICE, por sus siglas en inglés) [16], SCAN [11], entre otros.

En este artículo, se optó por utilizar este último algoritmo debido a tres principales razones. La primera razón está relacionada a que se desconocen completamente las

etiquetas y el número de grupos. La segunda razón está relacionada con la función de costo de SCAN que lo hace robusto a transformaciones de imagen.

El agrupamiento debe de ser robusto a la rotación; sin embargo, debido a que el tamaño de la trayectoria es relevante, el agrupamiento debe ser sensible a la escala. La transformación de desplazamiento no debe ser considerada, ya que todas las trayectorias están centradas en la imagen (Figura 1).

La tercera razón está relacionada a que el algoritmo calcula un valor, relacionado con la distancia de cada imagen al centro de su grupo, que es de gran utilidad para calcular el número mínimo de grupos. El algoritmo SCAN se compone de tres fases: aprendizaje auto-supervisado, agrupamiento y auto etiquetado [11].

Durante la primera etapa, el algoritmo aprende información relevante de una gran base de datos que contiene imágenes, ej. CIFAR-10 [17]. Después, SCAN realiza un agrupamiento de imágenes de trayectorias utilizando el algoritmo de los K vecinos más próximos y una función de costo que destruye la información de bajo nivel, ej. información pixel, según las transformaciones de imagen definidas.

Por último, SCAN realiza un auto-etiquetado considerando las imágenes de trayectoria mejor agrupadas para aumentar su precisión. En la Tabla 1 se puede encontrar la descripción del algoritmo SCAN en pseudocódigo. Basado en el número de grupos previamente propuesto en [10], el primer agrupamiento de SCAN se llevó a cabo con 6 grupos.

Tabla 1. Pseudocódigo del algoritmo SCAN. Agrupamiento Semántico por Adaptación de los Vecinos más Próximos (SCAN).

-
1. **Entrada:** Base de datos D , Grupo C , Tarea τ , Red Neuronal Φ_0 y Φ_η , Vecinos $N_D = \{\}$.
 2. Optimizar Φ_0 con la tarea τ . (Tarea de pretexto)
 3. **Para** $X_i \in D$, **hacer**
 4. $N_D \leftarrow N_D \cup N_{X_i}$, con $N_{X_i} = K$ vecindario muestra de $\Phi_0(X_i)$
 5. **Fin del ciclo**
 6. **Mientras** SCAN-pérdida decremente **hacer** (Agrupamiento)
 7. Actualizar Φ_η con SCAN-pérdida.
 8. **Fin de ciclo**
 9. **Mientras** longitud(Y) incremente **hacer** (Autoetiquetado)
 10. $Y \leftarrow (\Phi_\eta(D) > \text{umbral})$
 11. Actualizar Φ_η con pérdida de entropía cruzada, i.e. $H(\Phi_\eta(D), Y)$
 12. **Fin de ciclo**
 13. **Regresar:** $\Phi_\eta(D)$. (D es dividido entre los grupos C)
-

Sin embargo, al observar las imágenes peores agrupadas (es decir, que se localizan lejos del centro del grupo), se observó que sería necesario un grupo adicional. Entonces, se volvió a ejecutar el algoritmo con 7 grupos y notamos que todas las imágenes, incluyendo aquellas lejanas al centro de su grupo, estaban satisfactoriamente agrupadas (Figura 1).

Aquí es importante notar que es redundante continuar agrupando las imágenes con 8 grupos o más. Esto se debe a que los grupos adicionales podrían ser agrupados en

super grupos, de tal forma que el número mínimo de grupos (7, en este caso) es mantenido [10]. Salvo el número de grupos, ningún hiper-parámetro fue modificado de la versión original del código.

El algoritmo se implementó en un notebook Jupyter (versión JupyterLab 3.3.2) bajo Anaconda versión 3.21.5 ejecutando Python versión 3.8.10 y la librería Scikit-learn versión 1.0.2 [18], pandas versión 1.4.2, numpy versión 1.21.5 y matplotlib versión 3.6. Para el análisis estadístico se compararon los valores promedio de cada parámetro de movilidad en cada concentración de ketanserina contra el control en cada grupo y tiempo.

Los efectos de la ketanserina en cada subpoblación se examinaron mediante un análisis de varianza unidireccional (ANOVA) en un modelo de contrastes lineales (contrastos ortogonales). Los datos se presentan como $M \pm SEM$. Para todas las pruebas, se consideró significativo $p < 0.05$. Las ANOVA se realizaron utilizando el paquete R car versión 2.1-0 [19]. Todos los datos fueron analizados estadísticamente con el software R versión 3.4.3 [20].

3. Resultados

La opción de imprimir las coordenadas de cada espermatozoide del complemento CASA_bgm [13] proporcionó los datos de una manera complicada de seguir. Por lo tanto, se modificó el código del plugin. La nueva versión del complemento BGM permite obtener correctamente las coordenadas de cada espermatozoide detectado, junto con los datos cinemáticos para cada espermatozoide (este complemento se llamó "CASA_RA" y está disponible en DOI: 10.7910 / DVN / YI3N4Q).

Se construyeron dos archivos con los resultados del CASA_RA: uno para las coordenadas (llamado archivo "traj", que contiene 1,634,950 filas) y otro para los parámetros de movilidad (llamado "parámetros de movilidad", que contiene 20,805 filas, 11,394 en tiempo 0 y 9411 en tiempo 30; ambos dataset están disponibles en DOI: 10.7910 / DVN / JDJVB7).

La construcción de los archivos fue tediosa y consumió mucho tiempo; por lo tanto, se escribió un script en Python para automatizar el proceso y construir los archivos de manera iterativa (el script en Python, log_motility_data-6.py e instrucciones de uso están disponibles en doi: 10.7910 / DVN / YI3N4Q).

Los archivos de los parámetros de movilidad y de las trayectorias fueron tomados por otro script en Python para construir una instancia de la clase dataframe, la cual se nombró "master dataframe". En el master dataframe, las filas corresponden a los espermatozoides individuales; dos columnas contienen datos categóricos para las diferentes concentraciones de ketanserina y tiempo, respectivamente; ocho columnas contienen valores flotantes para cada uno de los ocho parámetros de movilidad evaluados; finalmente, una columna contiene las listas de coordenadas.

Un script de Python tomó las coordenadas en el master dataframe como entrada para construir la imagen de cada trayectoria y las guardó en un nuevo directorio (el script está disponible como un notebook Jupyter llamado "traj-ah-6-full-0.ipynb" en doi: 10.7910 / DVN / CBMKVA).

Las imágenes de trayectorias sirvieron como entrada para el algoritmo SCAN.

Tabla 2. Efecto de las diferentes concentraciones de ketanserina en las subpoblaciones, identificadas por algoritmos de aprendizaje de máquinas SCAN. Las mediciones de la movilidad espermática se realizaron inmediatamente después de poner ketanserina o después de 30 minutos de incubación. Los datos representan la *media* + SE.

Tiempo	Sub-población	Ketanserina (nM)	VCL (µm/s)	VAP (µm/s)	VSL (µm/s)	LIN (%)	STR (%)	WOB (%)	BCF (Hz)	Scan_values	n	
0	5	0	99.08 ± 1.36	61.23 ± 0.77	43.66 ± 0.74	0.46 ± 0.01	0.72 ± 0.01	0.63 ± 0.00	39.54 ± 0.24	0.65 ± 0.01	353	
		10	107.11 ± 1.74*	64.88 ± 0.96*	44.81 ± 0.90	0.44 ± 0.01	0.70 ± 0.01	0.62 ± 0.01	39.34 ± 0.27	0.65 ± 0.01	314	
		100	105.97 ± 2.18*	62.71 ± 1.06	43.65 ± 0.92	0.44 ± 0.01	0.71 ± 0.01	0.61 ± 0.01*	39.79 ± 0.27	0.63 ± 0.01	284	
		1000	111.67 ± 2.24*	67.87 ± 1.09*	49.62 ± 0.92*	0.48 ± 0.01	0.75 ± 0.01	0.63 ± 0.01	38.93 ± 0.27	0.69 ± 0.01*	294	
	6	0	103.39 ± 1.49	48.59 ± 0.68	28.06 ± 0.60	0.28 ± 0.01	0.59 ± 0.01	0.48 ± 0.00	38.31 ± 0.20	0.44 ± 0.01	593	
		10	116.62 ± 1.98*	54.05 ± 0.95*	31.04 ± 0.82*	0.28 ± 0.01	0.58 ± 0.01	0.47 ± 0.00	37.75 ± 0.24	0.47 ± 0.01*	512	
		100	113.31 ± 2.09*	52.69 ± 1.00*	30.10 ± 0.83	0.27 ± 0.01	0.59 ± 0.01	0.47 ± 0.00	38.46 ± 0.24	0.48 ± 0.01*	469	
		1000	129.44 ± 1.88*	58.95 ± 0.89*	31.77 ± 0.75*	0.25 ± 0.01*	0.55 ± 0.01*	0.46 ± 0.00*	38.13 ± 0.21	0.50 ± 0.01*	661	
	30	5	0	99.21 ± 1.73	65.31 ± 0.96	51.15 ± 0.93	0.54 ± 0.01	0.79 ± 0.01	0.67 ± 0.00	38.90 ± 0.27	0.70 ± 0.01	382
			10	102.36 ± 1.86	65.29 ± 0.99	45.87 ± 0.85*	0.48 ± 0.01*	0.72 ± 0.01*	0.65 ± 0.01*	39.05 ± 0.27	0.67 ± 0.01	303
			100	107.53 ± 1.95*	65.86 ± 1.06	45.91 ± 0.98*	0.45 ± 0.01*	0.71 ± 0.01*	0.62 ± 0.00*	39.40 ± 0.27	0.66 ± 0.01*	328
			1000	110.65 ± 1.75*	68.74 ± 0.92*	50.96 ± 0.85	0.48 ± 0.01*	0.75 ± 0.01*	0.64 ± 0.00*	38.61 ± 0.24	0.68 ± 0.01*	462
6		0	100.32 ± 2.81	49.95 ± 1.49	34.78 ± 1.30	0.35 ± 0.01	0.71 ± 0.01	0.50 ± 0.01	38.17 ± 0.39	0.41 ± 0.01	215	
		10	101.50 ± 2.99	50.89 ± 1.69	32.69 ± 1.38	0.33 ± 0.01	0.67 ± 0.02	0.50 ± 0.01	38.84 ± 0.43	0.44 ± 0.01*	179	
		100	104.75 ± 2.47	48.05 ± 1.26	32.29 ± 1.16	0.31 ± 0.01*	0.67 ± 0.01	0.46 ± 0.01*	39.30 ± 0.33*	0.42 ± 0.01	264	
		1000	109.21 ± 2.33*	51.88 ± 1.20	33.77 ± 1.03	0.32 ± 0.01*	0.66 ± 0.01*	0.48 ± 0.01*	38.00 ± 0.30	0.45 ± 0.01*	336	

*Indica diferencia significativa (P<0.05, ANOVA de una vía) vs el control en el mismo cluster. VCL, velocidad curvilínea; VAP, velocidad promedio; VSL, velocidad rectilínea; LIN, linealidad; STR, rectitud; WOB, bamboleo; BCF, frecuencia de batido; n, es el número de espermatozoides en cada cluster.

Después del agrupamiento, el script de Python creó una nueva columna con los identificadores de los grupos (subpoblaciones) para cada fila en el master dataframe original. Finalmente, la descripción estadística de las subpoblaciones de imágenes de trayectorias, en cada tratamiento con ketanserina y tiempo, se realizó en los subdataset de los parámetros de movilidad.

Después del agrupamiento, el script de Python creó una nueva columna con los identificadores de los grupos (subpoblaciones) para cada fila en el master dataframe original. Finalmente, la descripción estadística de las subpoblaciones de imágenes de trayectorias, en cada tratamiento con ketanserina y tiempo, se realizó en los subdataset de los parámetros de movilidad.

Estos subconjuntos se obtuvieron del master dataframe con base en los datos categóricos y los identificadores de subpoblaciones. El algoritmo SCAN permitió agrupar las imágenes de las trayectorias espermáticas en siete subpoblaciones, numeradas de 0 a 6 para cada tiempo de evaluación. Para obtener una representación gráfica de las imágenes de trayectorias en cada subpoblación, construimos gráficos "Pollock" (Figura 2). El agrupamiento de las imágenes en las subpoblaciones 5 y 6 corresponden a trayectorias largas curvadas.

Las trayectorias con curvas más cerradas, incluso formando bucles se observaron en la subpoblación 6. El porcentaje de espermatozoides en los grupos 0-6 en el tiempo 0 fue de 2.39, 19.94, 16.30, 17.21, 12.92, 11.16 y 20.04, respectivamente; mientras que en el tiempo 30 fue de 3.06, 25.59, 15.83, 16.48, 12.60, 15.77 y 10.63, respectivamente.

A continuación, se describirá el efecto de la ketanserina en los espermatozoides de las subpoblaciones 5 y 6 puesto que son aquellas que presentaron las velocidades más altas en los controles, lo cual es biológicamente relevante; sin embargo, los efectos de la ketanserina en el resto de las subpoblaciones se encuentran en la Tabla 2.

Notablemente en las subpoblaciones 5 y 6 también se observaron los valores más altos del índice de clasificación de SCAN (Fig. 3).

Los efectos de la ketanserina en las trayectorias de los espermatozoides en la subpoblación 5 y 6 se muestran en la Figura 3. Al tiempo 0 las velocidades de los controles fueron significativamente mayores en los tratamientos con altas concentraciones de ketanserina (VCL, VAP y VSL); mientras que la linealidad de las trayectorias (LIN y STR) no se modificó.

Sin embargo, a los 30 minutos se observó una disminución significativa en los valores de linealidad en los tratamientos con ketanserina, principalmente en la subpoblación 5. Los índices de SCAN aumentaron en el tiempo 0, pero tienden a disminuir al tiempo 30 (Tabla 2). Gráficamente estos cambios se aprecian como trayectorias más curvadas (Fig. 3).

Al tiempo 0 los índices promedio de SCAN en la subpoblación 5 aumentó significativamente en la concentración más alta de ketanserina; pero aumentó significativamente en todas las concentraciones de ketanserina en la subpoblación 6. Al tiempo 30 los valores de SCAN disminuyeron significativamente en los tratamientos con ketanserina en la subpoblación 5, pero aumentaron significativamente en la subpoblación 6 (Fig. 3).

4. Conclusiones y trabajos a futuro

Hasta ahora, la identificación estadística de subpoblaciones cinemáticas se desarrolló en función de los parámetros de movilidad, obtenidos de los sistemas CASA. Una vez identificadas las subpoblaciones cinemáticas, algunos autores proporcionan imágenes representativas de las trayectorias seguidas por uno o unos pocos espermatozoides en cada subpoblación.

Hasta donde sabemos, este es el primer trabajo en el que se reconstruye la trayectoria de cada espermatozoide evaluado en un sistema CASA, para utilizarlas como entrada para un algoritmo de agrupamiento de aprendizaje automático; identificando y describiendo así subpoblaciones cinemáticas.

Los patrones de movilidad mencionados son la representación gráfica de las trayectorias de los espermatozoides observadas durante un período determinado. Aunque la dimensión tiempo estuvo ausente en el análisis de las imágenes de trayectorias, se recuperó parcialmente cuando los valores de los parámetros de movilidad se asociaron con las imágenes de trayectorias. Anteriormente, se ha informado sobre la importancia de algunos patrones de movilidad en la fertilización [21].

En contraste con la noción convencional de que los espermatozoides fértiles siempre son vigorosos, móviles y más lineales, los resultados en toros han demostrado que los patrones sinuosos están asociados con la fertilidad [21]. En nuestro trabajo, encontramos subpoblaciones caracterizadas por la presencia de trayectorias sinuosas (subpoblaciones 5 y 6) y no lineales.

Esto concuerda con el hecho de que porcentajes más significativos de espermatozoides rápidos y no lineales parecen tener una mayor capacidad de fertilización [22]. El manejo estadístico de los datos impacta en la predicción de la fertilidad. Por ejemplo, en el análisis de subpoblaciones de toros asociado con el

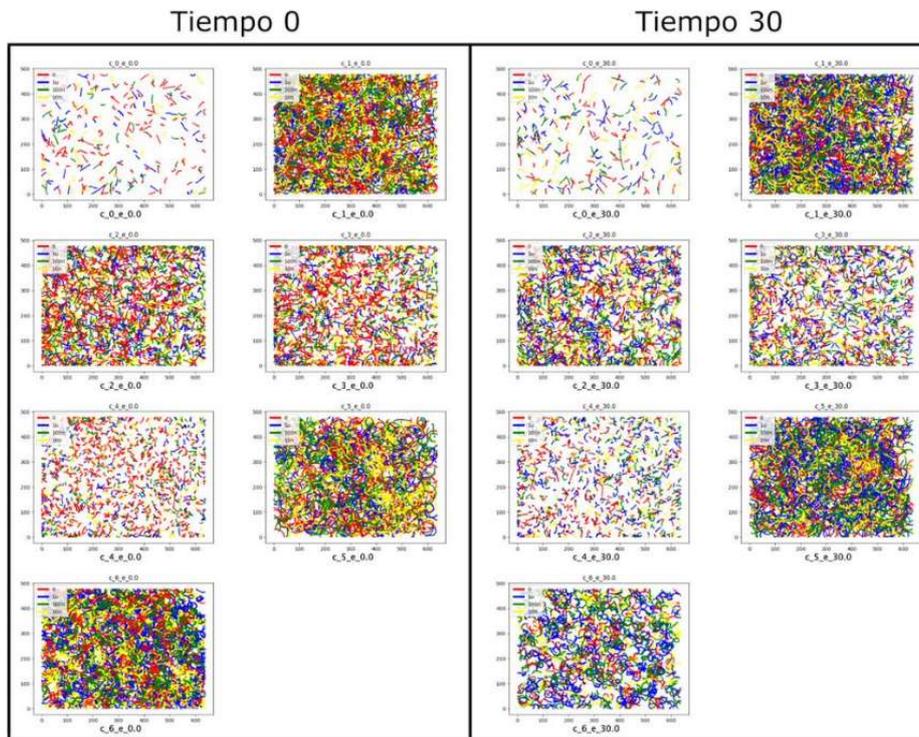


Fig. 2. Las imágenes de las trayectorias en cada tratamiento con ketanserina se encuentran en todos los grupos obtenidos por el algoritmo SCAN. En cada gráfico de Pollock se dibujó la trayectoria de los espermatozoides individuales en cada subpoblación. Las trayectorias en cada subpoblación lucen muy similares; mientras que la distribución de las trayectorias, coloreadas por tratamiento parecen distribuirse aleatoriamente. Las subpoblaciones 5 y 6 presentaron trayectorias largas y curvadas.

método de agrupamiento se puede considerar efectivo para predecir una mayor fertilidad [22].

Sin embargo, en cerdos, se agrupan los eyaculados, y el promedio de cada una de las variables cinemáticas tiene una capacidad predictiva limitada con respecto al tamaño de la camada y las variables de fertilidad [23]. Estudios similares a los previamente descritos, podrían aprovechar el método computacional que se describe aquí.

En este trabajo, se utilizaron espermatozoides de dosis de semen listas para inseminar. Por lo tanto, el semen se encontraba diluido, pero los espermatozoides no estaban lavados; por lo que, posiblemente la serotonina, entre otras sustancias presentes en el medio, podrían afectar la movilidad espermática.

La presencia de serotonina en el semen está respaldada por el hecho de que la serotonina está presente en las vesículas seminales [24] y en la próstata [25], y la emisión de espermatozoides desde el testículo va acompañada de líquido seminal de las vesículas seminales y la próstata. En este trabajo, las subpoblaciones se formaron con las imágenes de trayectorias y se describieron cuantitativamente con los valores de los parámetros de movilidad asociados con las trayectorias.

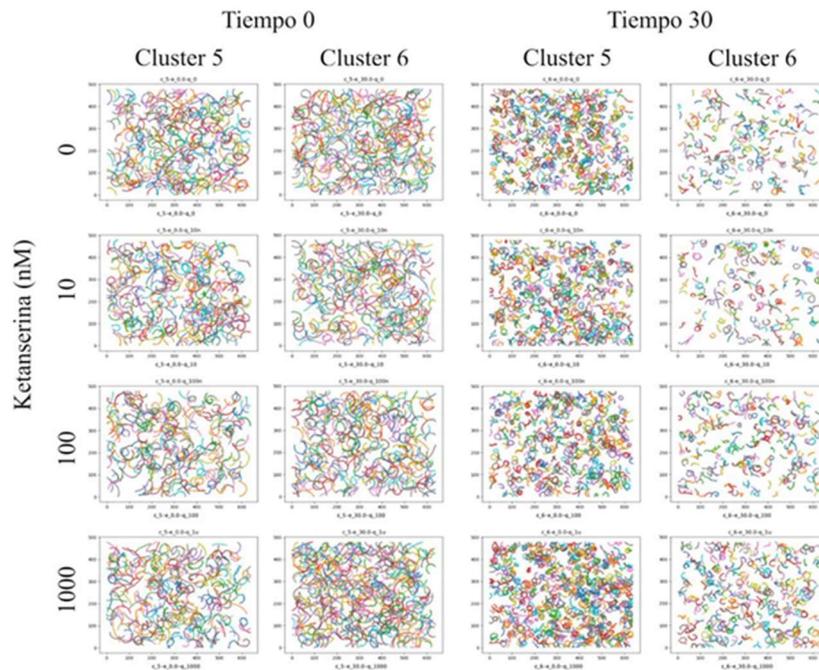


Fig. 3. La ketanserina indujo cambios en las trayectorias espermáticas. El dataset de imágenes de trayectorias sirvió como entrada para el algoritmo SCAN. Se identificaron 7 subpoblaciones. Los identificadores de tiempo de incubación y tratamiento con ketanserina se utilizaron para obtener los subdataset en cada subpoblación. Los gráficos de Pollock indican los cambios en las trayectorias espermáticas en las diferentes concentraciones de ketanserina al tiempo 0 y 30. Los espermatozoides en las subpoblaciones 5 y 6 presentaron las velocidades más altas y las trayectorias más curvas. La curvatura de las trayectorias tiende a aumentar después de 30 minutos de incubación con ketanserina. Los colores de las trayectorias son aleatorios.

Aquí, se proporciona un modelo computacional para discriminar imágenes de trayectorias, sin el sesgo inducido por el técnico. Nuestro modelo computacional sirvió para obtener una descripción detallada de las subpoblaciones cinemáticas después de la exposición a la ketanserina.

Este enfoque novedoso para manejar datos de movilidad podría proporcionar nuevas ideas sobre el papel de la movilidad espermática en la producción animal y la investigación básica sobre la regulación fisiológica de la movilidad espermática.

Las imágenes de las trayectorias tienen diferencias sutiles entre sí; estas diferencias están ocultas en los parámetros promedio de movilidad en cada subpoblación cinemática. Por otro lado, la forma de las trayectorias es diferente en longitud, grado de curvatura de la trayectoria completa, amplitud, frecuencia y forma.

El metabolismo de la ketanserina podría ayudar a explicar los resultados del efecto en VSL después de 30 minutos de incubación. La ketanserina en plasma decae triexponencialmente con vidas medias secuenciales de 0.13, 2 y 14.3 horas [26], lo cual concuerda con los cambios en los efectos en VCL y VSL observados después de 30 minutos en nuestro trabajo.

Por otro lado, el rápido aumento de las velocidades de los espermatozoides en los grupos 5 y 6 parece intrigante, puesto que se esperaba un efecto opuesto. Sin embargo, si la serotonina estuviera presente en el eyaculado y otros receptores 5-HT también están presentes en el espermatozoide de cerdo, entonces el antagonismo en los receptores 5-HT2 por la ketanserina favorecería la señalización por estos otros receptores.

Los diferentes resultados entre el trabajo de [10] y nuestro estudio podrían explicarse por factores como la especie, el uso de medios capacitantes durante la incubación, el tiempo de incubación antes de realizar las mediciones de movilidad, el antagonista distinto del receptor de serotonina utilizado entre trabajos y la estrategia de análisis de datos; es decir, el uso de valores promedio de parámetros de movilidad en [10] y el uso de un algoritmo de aprendizaje de máquinas, especializado en el agrupamiento de imágenes, utilizado en este trabajo.

En conclusión, el modelo computacional de agrupamiento de imágenes de trayectorias, desarrollado en este trabajo, proporciona un enfoque muy sutil para identificar el efecto de la ketanserina en los patrones de movilidad. La identificación de subpoblaciones cinemáticas siguiendo nuestro enfoque podría ayudar a identificar proporciones de espermatozoides relevantes para las técnicas de reproducción artificial.

Nuestro enfoque podría adoptarse para datos de otras especies (por ejemplo, humanos) y para reevaluar datos de trabajos anteriores, donde únicamente se utilizaron los parámetros de movilidad para identificar subpoblaciones cinemáticas (solamente se necesitarían los valores de coordenadas y parámetros de movilidad).

Agradecimientos. Este proyecto se desarrolló gracias al apoyo DGAPA-PAPIIT de la UNAM número IT201021.

Referencias

1. Martínez-Pastor, F.: What is the importance of sperm subpopulations? *Animal Reproduction Science*, vol. 246, p. 106844 (2022) doi: 10.1016/j.anireprosci.2021.106844
2. Ramón, M., Martínez-Pastor, F.: Implementation of novel statistical procedures and other advanced approaches to improve analysis of CASA data. *Reproduction, Fertility and Development*, vol. 30, no. 6, pp. 860–866 (2018) doi: 10.1071/RD17479
3. Ibănescu, I., Leiding, C., Bollwein, H.: Cluster analysis reveals seasonal variation of sperm subpopulations in extended boar semen. *Journal of Reproduction and Development*, vol. 64, no. 1, pp. 33–39 (2018) doi: 10.1262/jrd.2017-083
4. Henning, H., Petrunkina, A. M., Harrison, R. A. P., Waberski, D.: Cluster analysis reveals a binary effect of storage on boar sperm motility function. *Reproduction, Fertility and Development*, vol. 26, no. 5, pp. 623–632 (2014) doi: 10.1071/RD13113
5. Goodson, S. G., Zhang, Z., Tsuruta, J. K., Wang, W., O'Brien, D. A.: Classification of mouse sperm motility patterns using an automated multiclass support vector machines model. *Biology of reproduction*, vol. 84, no. 6, pp. 1207–1215 (2011) doi: 10.1095/biolreprod.110.088989
6. Gacem, S., Valverde, A., Catalán, J., Yáñez Ortiz, I., Soler, C., Miró, J.: A new approach of sperm motility subpopulation structure in donkey and horse. *Frontiers in Veterinary Science*, vol. 8 (2021) doi: 10.3389/fvets.2021.651477
7. Jimenez-Trejo, F., Tapia-Rodríguez, M., Cerbon, M., Kuhn, D. M., Manjarrez-Gutierrez, G., Mendoza-Rodríguez, C. A., Picazo, O.: Evidence of 5-HT components in human sperm:

- implications for protein tyrosine phosphorylation and the physiology of motility. *Reproduction*, vol. 144, no. 6, pp. 677–685 (2012) doi: 10.1530/REP-12-0145
8. Fujinoki, M.: Serotonin-enhanced hyperactivation of hamster sperm. *Reproduction*, vol. 142, no. 2, pp. 255–266 (2011) doi: 10.1530/REP-11-0074
 9. Sakamoto, C., Fujinoki, M., Kitazawa, M., Obayashi, S.: Serotonergic signals enhanced hamster sperm hyperactivation. *Journal of Reproduction and Development*, vol. 67, no. 4, pp. 241–250 (2021) doi: 10.1262/jrd.2020-108
 10. Rodríguez-Martínez, E. A., Rivas, C. U., Ayala, M. E., Blanco-Rodríguez, R., Juárez, N., Hernández-Vargas, E. A., Aragón, A.: A new computational approach, based on images trajectories, to identify the subjacent heterogeneity of sperm to the effects of ketanserin. *Cytometry Part A* (2023) doi: 10.1002/cyto.a.24732
 11. Van-Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., Van-Gool, L.: Scan: learning to classify images without labels. In: 16th European Conference on Computer Vision, pp. 268–285 (2020) doi: 10.1007/978-3-030-58607-2_16
 12. Rivas, C., Ayala, M. E., Aragón, A.: Effect of various pH levels on sperm kinetic parameters of boars. *South African Journal of Animal Science*, vol. 52, no. 5, pp. 693–704 (2022) doi: 10.4314/sajas.v52i5.13
 13. Giaretta, E., Munerato, M., Yeste, M., Galeati, G., Spinaci, M., Tamanini, C., Mari, G., Bucci, D.: Implementing an open-access CASA software for the assessment of stallion sperm motility: Relationship with other sperm quality parameters. *Animal Reproduction Science*, vol. 176, pp. 11–19 (2017) doi: 10.1016/j.anireprosci.2016.11.003
 14. Lundh, F., Clark, J. A.: Image module (2022) pillow.readthedocs.io/en/stable/reference/Image.html
 15. Mahon, L., Lukasiewicz, T.: Selective pseudo-label clustering. In: *Advances in Artificial Intelligence: 44th German Conference on AI, Virtual Event*, pp. 158–178 (2021) doi: 10.1007/978-3-030-87626-5_12
 16. Niu, C., Shan, H., Wang, G.: Spice: Semantic pseudo-labeling for image clustering. *IEEE Transactions on Image Processing*, vol. 31, pp. 7264–7278 (2022) doi: 10.1109/TIP.2022.3221290
 17. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images (2009)
 18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.: Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830 (2011)
 19. Fox, J., Weisberg, S.: *An R companion to applied regression*. Sage Publications (2011)
 20. R core team: *R: A language and environment for statistical computing* (2015) www.R-project.org/
 21. Nagata, M. P. B., Endo, K., Ogata, K., Yamanaka, K., Egashira, J., Katafuchi, N., Yamanouchi, T., Matsuda, H., Goto, Y., Sakatani, M., Hojo, T., Nishizono, H., Yotsushima, K., Takenouchi, N., Hashiyada, Y., Yamashita, K.: Live births from artificial insemination of microfluidic-sorted bovine spermatozoa characterized by trajectories correlated with fertility. In: *Proceedings of the National Academy of Sciences*, vol. 115, no. 14, pp. E3087–E3096 (2018) doi: 10.1073/pnas.1717974115
 22. Tsunokawa-Hidalgo, M. M., Marques-de-Almeida, A. B., Zito-de-Moraes, F. L., Palhaci-Marubayashi, R. Y., Ferreira-de-Souza, F., Rigo-Barreiros, T. R., Mello-Martins, M. I.: Sperm subpopulations influence the pregnancy rates in cattle. *Reproduction in Domestic Animals*, vol. 56, no. 8, pp. 1117–1127 (2021) doi: 10.1111/rda.13955
 23. Barquero, V., Roldan, E. R. S., Soler, C., Vargas-Leitón, B., Sevilla, F., Camacho, M., Valverde, A.: Relationship between fertility traits and kinematics in clusters of boar ejaculates. *Biology*, vol. 10, no. 7, p. 595 (2021) doi: 10.3390/biology10070595.
 24. Hsieh, J. T., Liu, S. P., Hsieh, C. H., Lai, M. K., Cheng, J. T.: An ex vivo evaluation of regulatory role of biogenic amines in rat seminal vesicle after pharmacological

Eder Alejandro Rodríguez Martínez, Cindy Ursula Rivas Arsaluz, Andrés Aragón Martínez

- manipulation. *Life Sciences*, vol. 63, no. 15, pp. PL221–PL229 (1998) doi: 10.1016/S0024-3205(98)00403-2
25. Mota, P., Barbosa-Martins, J., Moura, R. S., Lima, E., Miranda, A., Correia-Pinto, J., Carvalho-Dias, E.: Effects of testosterone replacement on serotonin levels in the prostate and plasma in a murine model of hypogonadism. *Scientific Reports*, vol. 10, no. 1, p. 14688 (2020) doi: 10.1038/s41598-020-71718-z
 26. Persson, B., Heykants, J., Hedner, T.: Clinical pharmacokinetics of ketanserine. *Clinical Pharmacokinetics*, vol. 20, no. 4, pp. 263–279 (1991) doi: 10.2165/00003088-199120040-00002