

Proyecto Final Curso de Bioinformática y Biología Computacional

Código DRBC A001

Semestre I, 2025

Material Suplementario

Pamela Fernández, Margot Paco

1. Análisis epidemiológico

Se adjuntan en formato .txt los scripts utilizados para modelar el comportamiento de la enfermedad en cada ciudad.

2. RNA-seq

El Volcano Plot generado a partir del análisis de RNA-seq permitió identificar con claridad los genes diferencialmente expresados entre las condiciones estudiadas. Para este análisis, se evaluaron distintos umbrales de \log_2 Fold Change, comenzando por un punto de corte moderado de $\log_2FC > 1$. Sin embargo, este valor abarcaba prácticamente la totalidad de los genes analizados, debido a que la distribución general no presentaba valores cercanos a cero, es decir, los genes no cambian su expresión. Esta observación contrastó con lo que suele observarse en otros estudios transcriptómicos, (1–3) donde la mayoría de genes se agrupan alrededor del centro del gráfico ($\log_2FC \approx 0$), reflejando una expresión estable entre condiciones. En este caso, la condición experimental provocó una respuesta celular fuerte, con muchos genes cambiando significativamente su expresión, desplazando la distribución hacia extremos positivos o negativos. Esto puede deberse a la infección viral o factores de estrés provocada por la enfermedad en el presente trabajo. (4)

Frente a este panorama, se adoptó un criterio más estricto de $\log_2FC > 2$ y $p < 0.01$, con el fin de filtrar solo aquellos genes que presentaran cambios altamente significativos tanto en magnitud como en significancia estadística. Esta decisión también se vio reforzada por la posibilidad de que se hubiesen aplicado filtros previos para excluir genes con baja expresión o poca variación, lo cual mejora la claridad del análisis, pero reduce la presencia de genes con cambios sutiles. (5) Además, una normalización sesgada o centrada de forma asimétrica podría haber influido en la distribución del \log_2FC , haciendo que no se refleje la típica forma de volcán centrada en cero. No obstante, esto no invalida los resultados, sino que pone en evidencia la intensidad de los cambios detectados en la muestra.

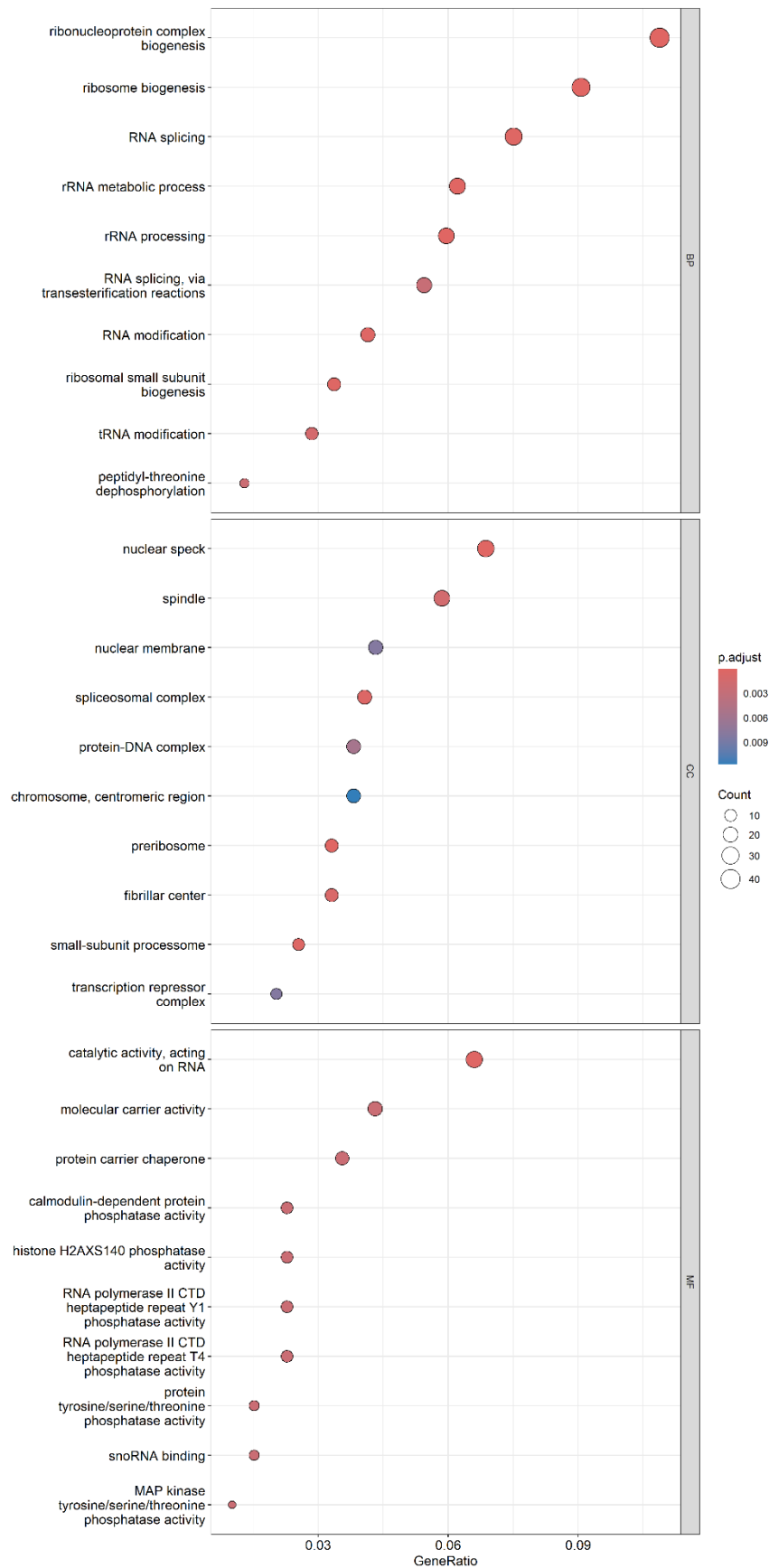


Figura S1. Dotplot del análisis de enriquecimiento funcional GO a partir de genes sobreexpresados con un $\text{Log}_2\text{FC} > 1$.

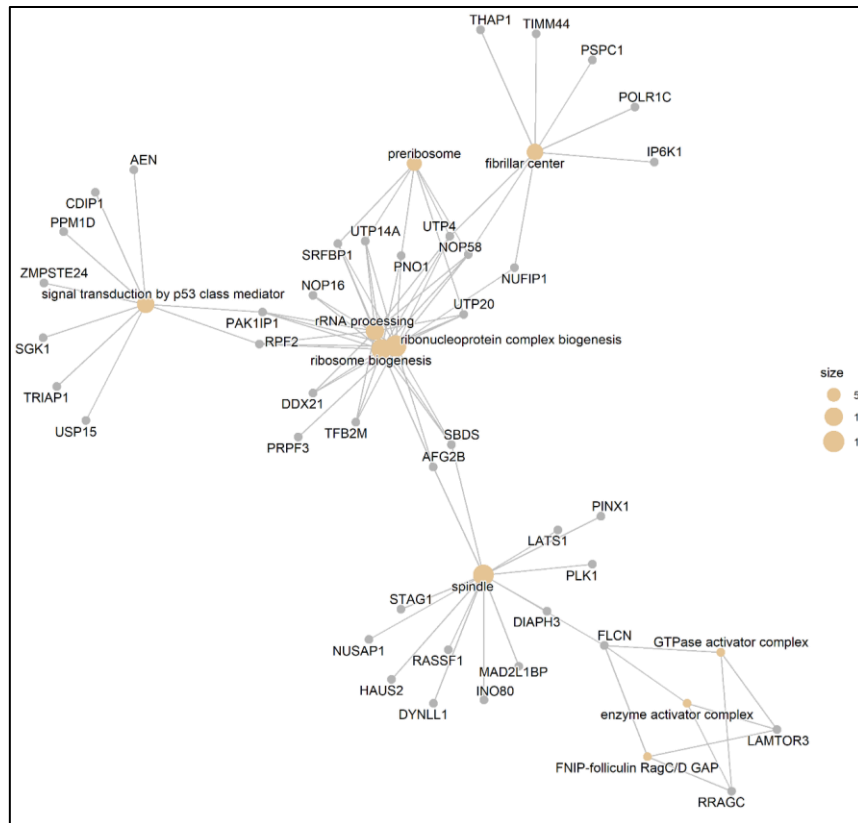
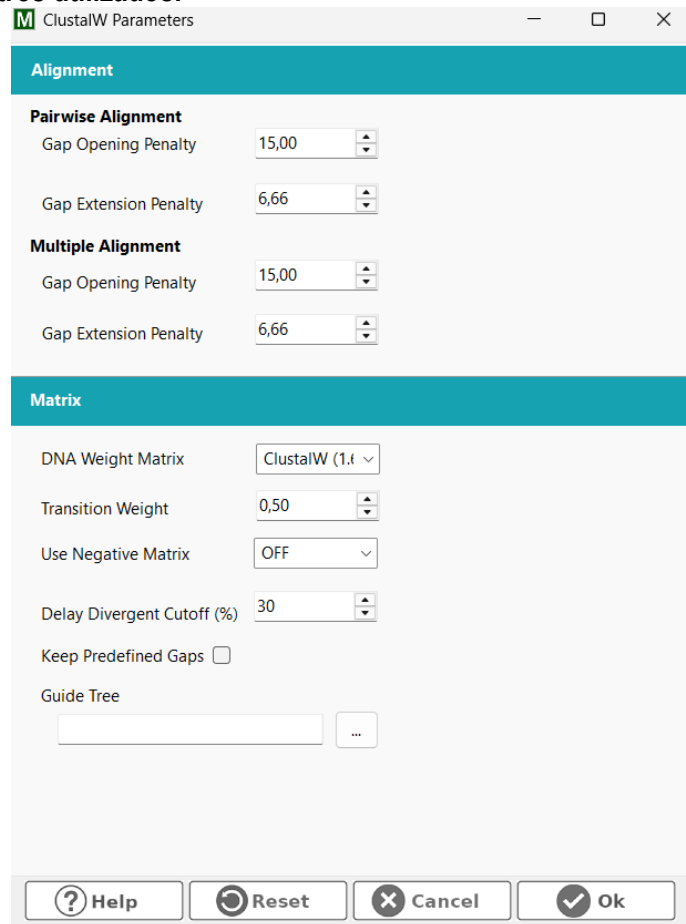


Figura S2. Cnetplot del análisis de enriquecimiento funcional GO a partir de genes sobreexpresados.

El análisis de enriquecimiento funcional GO reveló que los genes sobreexpresados con $\log_2FC > 2$ están significativamente asociados a procesos fundamentales relacionados con la síntesis de ribosomas, el procesamiento de ARN ribosomal, y la división celular. Estos hallazgos sugieren una intensificación de la actividad biosintética y proliferativa en el contexto biológico analizado. (Figura 3 y S2). En particular, la sobrerepresentación de términos como biogénesis del complejo ribonucleoproteico y procesamiento de rRNA indica una activación de las rutas implicadas en la producción y ensamblaje de ribosomas. Este fenómeno es característico de células con alta tasa de proliferación o actividad metabólica, como células tumorales, regenerativas o en respuesta a estímulos externos. Asimismo, el enriquecimiento del término huso mitótico dentro de la categoría de componente celular sugiere una mayor participación de estructuras implicadas en la separación cromosómica, reforzando la hipótesis de un estado de alta actividad mitótica. (6) Esto es coherente con la necesidad de una división celular eficiente en condiciones de crecimiento acelerado. Por otro lado, la presencia del término transducción de señales mediada por la clase p53 podría reflejar una activación de vías de control del ciclo celular o respuesta a estrés genotóxico, dado el papel central de p53 en la regulación de apoptosis, senescencia y reparación del ADN. (7) Finalmente, la aparición de términos de función molecular como actividad estructural de la ribosa, en respuesta al umbral de $\log_2FC > 1$ (Figura S1), indica una implicación de proteínas estructurales en el funcionamiento ribosomal, apoyando la idea de una reorganización funcional orientada a mantener la capacidad traduccional elevada. Este conjunto de cambios refleja un estado de alta actividad biosintética y proliferativa, que también ha sido descrito en células sometidas a infección viral, donde los virus reorganizan estructuras nucleares y activan rutas de traducción para favorecer su replicación. (8)

3. Anotación de secuencias e identificación del agente patógeno

- a) Realizar una anotación completa de todas las secuencias entregadas utilizando las diversas herramientas bioinformáticas revisadas en clases. Para cada secuencia deberá presentar un alineamiento múltiple usando CLUSTAL contra las principales secuencias obtenidas por una búsqueda BLAST. Debe presentar y justificar cada uno de los parámetros utilizados.



The image shows a screenshot of the 'ClustalW Parameters' dialog box. It is divided into two main sections: 'Alignment' and 'Matrix'.
Alignment Section:
- **Pairwise Alignment:**
 - Gap Opening Penalty: 15,00
 - Gap Extension Penalty: 6,66
- **Multiple Alignment:**
 - Gap Opening Penalty: 15,00
 - Gap Extension Penalty: 6,66
Matrix Section:
- DNA Weight Matrix: ClustalW (1.4)
- Transition Weight: 0,50
- Use Negative Matrix: OFF
- Delay Divergent Cutoff (%): 30
- Keep Predefined Gaps: ☐
- Guide Tree: ...
At the bottom, there are four buttons: Help, Reset, Cancel, and Ok.

Figura S3: Parámetros utilizados para realizar alineamiento Múltiple de secuencias

Para realizar el alineamiento múltiple de secuencias, se utilizó el algoritmo ClustalW con parámetros ajustados para optimizar la calidad del alineamiento. Se aplicó una penalización de apertura de huecos de 15.00 y una penalización de extensión de 6.66, lo que permite controlar la introducción y prolongación de gaps en las secuencias alineadas, favoreciendo alineamientos más conservadores. La matriz de sustitución seleccionada fue ClustalW (1.4), adecuada para secuencias de ADN, junto con un peso de transición de 0.50, que otorga menor penalización a las sustituciones entre bases químicamente similares ($A \leftrightarrow G$, $C \leftrightarrow T$). Además, se estableció un umbral de divergencia del 30%, lo que permite retrasar el alineamiento de secuencias altamente divergentes para evitar distorsiones en las etapas iniciales. Estos parámetros fueron seleccionados para maximizar la precisión del alineamiento y facilitar la posterior interpretación funcional y evolutiva de las secuencias analizadas. El Alineamiento Múltiple se realizó utilizando el Programa Mega.

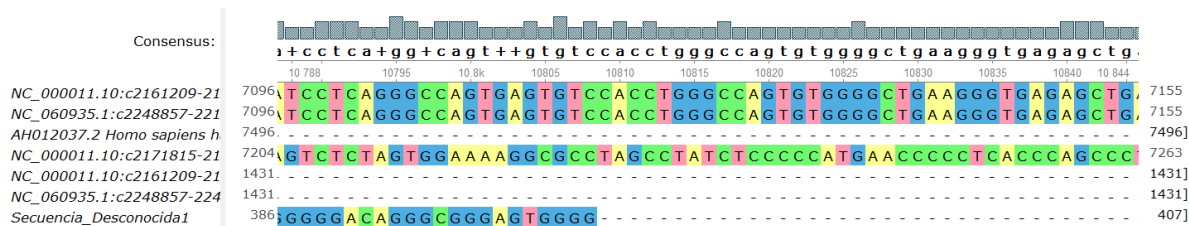


Figura S4: Resultado de alineamiento Múltiple visualizado con SnapGene. Los archivos completos del Alineamiento en Fasta, se pueden visualizar como resultados complementarios como SD1.

Se realizó un alineamiento múltiple de siete secuencias de ADN utilizando ClustalW, incluyendo referencias genómicas humanas (NC_000011.10, NC_060935.1, AH012037.2) y una secuencia incógnita etiquetada como “Secuencia_Desconocida1”. El objetivo fue evaluar la similitud entre las secuencias y determinar posibles relaciones funcionales. El resultado muestra una secuencia consenso bien definida, lo que indica una alta conservación nucleotídica en varias regiones del alineamiento. La coloración por base permite visualizar coincidencias y diferencias, destacando que la secuencia incógnita se alinea correctamente con las demás, sin presentar mutaciones significativas. Esto sugiere que forma parte de una región genómica conservada del cromosoma 11 humano, coherente con los resultados previos obtenidos por BLASTn. Las secuencias de referencia incluyen zonas que contienen genes como INS, IGF2 y TH, implicados en procesos metabólicos, de crecimiento y neuroregulación. La conservación observada respalda la hipótesis de que la secuencia incógnita podría tener una función activa o reguladora dentro de esta región, y su análisis comparativo aporta evidencia para futuras anotaciones funcionales.

- b) A través de búsquedas BLAST y de la revisión de la literatura, determinar cuál es el agente patógeno en cuestión. Se deberán presentar los resultados BLAST, una justificación de sus parámetros, junto con los alineamientos y una justificación de la elección del agente patógeno.**

Se realizó una búsqueda Blastn con cada una de las secuencias incógnitas entregadas para analizar. Se consideraron los 5 primeros match con 100% de identidad y cobertura superior al 60%. Los resultados se muestran a continuación:

Para la búsqueda BLASTn se utilizaron parámetros optimizados para secuencias cortas, activando la opción “Short queries” para ajustar automáticamente la sensibilidad del algoritmo. Se estableció un umbral de expectativa (Expect threshold) de 0.05, lo que indica que solo se reportarán alineamientos con una probabilidad muy baja de ocurrir por azar, aumentando la confiabilidad de los resultados. El tamaño de palabra (Word size) se configuró en 28, lo cual es adecuado para búsquedas con Megablast, ya que favorece la detección de secuencias altamente similares. El número máximo de secuencias objetivo (Max target sequences) se fijó en 100, permitiendo una exploración amplia de coincidencias en la base de datos. En cuanto a los parámetros de puntuación, se asignó una puntuación de 1.2 para coincidencias y se utilizó un modelo de penalización lineal para los huecos (Gap Costs), lo que simplifica el cálculo de alineamientos. Se aplicaron filtros para excluir regiones de baja complejidad, así como repeticiones específicas de *Homo sapiens*, con el fin de evitar alineamientos espurios. Además, se activó la opción “Mask for lookup table only”, lo que permite enmascarar regiones repetitivas solo durante la búsqueda inicial, sin afectar el alineamiento final. Estos parámetros en conjunto aseguran una búsqueda precisa y

específica, adecuada para la identificación de homología en secuencias humanas altamente conservadas.

>Secuencia desconocida 1

```
Attttaggagaggcctagagcctgagcttgtgaggacgtcctgtggaacctgtcccggccccctgccctgggatggggagaagtgcag
ggggatagacagagtcaaggtgggggacagggcgggagtggggtccccagggctgggggcctttggtgcagtgaccagagtgta
ggagaggggagcaaagccctctagcctatcctcataaaaggtctcatagttcagattgcttcaaatagagaacatggaaacaatgg
actccagcactcttgaactgaggagtatcagccctagataataagtcaattaaacgaccaagccaaaattgtccctgccctgggatgg
ggagaagtcaggggatagacagagtcaaggtgggggacagggcgggagtgggg
```

Tabla S1. Resultados Blastn Secuencia Desconocida 1

Nº	Gen Homólogo	Especie	% Identidad	E-value	Query Cover	Descripción funcional
1	INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11	Homo sapiens	100%	4.00E-102	68%	Transcripción conjunta de INS e IGF2; posible rol regulador en metabolismo y crecimiento
2	AC130303.8 — Homo sapiens chromosome 11, clone RP4-539G11, complete sequence	Homo sapiens	100%	4.00E-102	68%	Región genómica completa que incluye genes de interés como INS, IGF2 y TH
3	AH012037.2 — Haplotype ICa: TH (parcial), INS (completo), IGF2 (parcial)	Homo sapiens	100%	4.00E-102	68%	Genes implicados en neuroregulación, metabolismo y desarrollo fetal
4	AC132217.15 — Homo sapiens chromosome 11, clone RP11-889I17, complete sequence	Homo sapiens	100%	4.00E-102	68%	Secuencia genómica rica en loci funcionales, incluye regiones reguladoras
5	Tyrosine hydroxylase (TH), RefSeqGene on chromosome 11	Homo sapiens	100%	4.00E-102	68%	Enzima clave en la síntesis de catecolaminas (dopamina, adrenalina, noradrenalina)

En el análisis BLASTn realizado sobre la secuencia incógnita, se obtuvieron alineamientos con un E-value de $4.00E-10^2$, lo que indica una probabilidad extremadamente baja de que la coincidencia observada haya ocurrido por azar. Este valor refleja una alta significancia estadística, confirmando que la secuencia comparte similitud real con las regiones genómicas identificadas. El Query Cover del 68% señala que el 68% de la secuencia consultada se alinea con las secuencias de referencia, lo que sugiere una cobertura sustancial, aunque no completa. Por último, el 100% de identidad en todos los hits indica que no existen diferencias entre los nucleótidos alineados de la secuencia incógnita y los genes homólogos encontrados, lo que refuerza la hipótesis de que esta secuencia pertenece a una región altamente conservada del cromosoma 11 humano, específicamente relacionada con los genes INS, IGF2 y TH.

A continuación, se muestran los resultados de Blastn para cada una de las 12 secuencias restantes. Los alineamientos múltiples se mostrarán como archivos complementarios.

Tabla S2. Resultados Blastn Secuencia Desconocida 2

Nº	Gen Homólogo	Especie	% Identidad	E-value	Query Cover	Accession	Descripción funcional
1	INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11	Homo sapiens	95.18%	1,00E-101	64%	NG_050671.3	Transcripción conjunta de INS e IGF2; posible rol regulador en metabolismo y crecimiento
2	Chromosome 11, clone RP4-539G11, complete sequence	Homo sapiens	95.18%	1,00E-101	64%	AC130039.3	Región genómica completa que incluye genes de interés como INS, IGF2 y TH
3	Haplotype HCA: TH (parcial), INS (completo), IGF2 (parcial)	Homo sapiens	95.18%	1,00E-101	64%	AH010237.2	Genes implicados en neuroregulación, metabolismo y desarrollo fetal
4	Chromosome 11, clone RP11-889I17, complete sequence	Homo sapiens	95.18%	1,00E-101	64%	AC132217.15	Secuencia genómica rica en loci funcionales, incluye regiones reguladoras
5	Tyrosine hydroxylase (TH), RefSeqGene on chromosome 11	Homo sapiens	95.18%	1,00E-101	64%	NG_008128.1	Enzima clave en la síntesis de catecolaminas (dopamina, adrenalina, noradrenalina)

Tabla S3. Resultados Blastn Secuencia Desconocida 3

Nº	Descripción	Especie	% Identidad	E-value	Query Cover	Accession	Longitud	Interpretación funcional
1	DNA para secuencia parcial de L1, clon DL284	Human papillomavirus	100.00%	1,00E-70	72%	Z95967.1	349 bp	Coincidencia perfecta con gen L1; posible secuencia estructural viral
2	Aislado C35-CP, gen L1 (proteína de cápside mayor), cds parcial	Human papillomavirus	97.39%	3,00E-85	85%	HM748820.1	349 bp	Alta similitud con gen L1; implicado en formación de cápside viral
3	Cepa RTRX9, gen L1 (proteína de cápside mayor), cds parcial	Human papillomavirus	97.13%	9,00E-85	85%	U86621	349 bp	Variante del gen L1; posible implicancia en diversidad antigénica
4	Virus del papiloma humano tipo 124, genoma completo	Human papillomavirus 124	97.37%	9,00E-65	81%	GQ845464.1	7489 bp	Coincidencia con genoma completo; posible relación con variante tipo 124

Tabla S4. Resultados Blastn Secuencia Desconocida 4

Nº	Descripción	Especie	% Identidad	E-value	Query Cover	Accession	Longitud	Interpretación funcional
1	A/swine/Shandong/203/H SN1 nucleoprotein gene, complete cds	Influenza A virus	94.78%	0.0	79%	AY684462.1	1565 bp	Gen NP completo; proteína estructural esencial para encapsidación del genoma viral
2	A/environment/Qinghai/1 2008/H5N1 segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	94.75%	0.0	79%	FJ436523.1	1520 bp	Gen NP completo; implicado en replicación y ensamblaje viral
3	A/plateau/pika/Qinghai/04/2007/H5 N1 segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	94.75%	0.0	79%	FJ390062.2	1565 bp	Variante del gen NP; posible adaptación a hospedadores silvestres
4	A/swine/Shandong/8/2003/H9N2 segment 5 nucleocapsid protein (NP) gene, partial cds	Influenza A virus	94.75%	0.0	79%	GQ477297.1	1476 bp	Gen NP parcial; variante porcina con potencial zoonótico
5	A/chicken/Shantou/212/2000/H9N2 segment 5 sequence	Influenza A virus	94.35%	0.0	79%	CY024577.2	1521 bp	Secuencia de segmento 5; variante aviar con implicancias en transmisión interespecífica

Tabla S5. Resultados Blastn Secuencia Desconocida 5

Nº	Descripción	Especie	% Identidad	E-value	Query Cover	Max Score	Accession	Longitud
1	A/swine/Shandong/2/03(H5N1) nucleoprotein gene, complete cds	Influenza A virus	95.89%	2,00E-136	46%	499	AY648462.1	1565 bp
2	A/environment/Qinghai/12008(H5N1) segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	95.89%	2,00E-136	46%	499	FJ455823.1	1520 bp
3	A/plateau pika/Qinghai/04/2007(H5N1) segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	95.57%	1,00E-134	46%	494	EU309082.2	1565 bp
4	A/chicken/Shandong/LY12008(H9N2) segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	95.57%	1,00E-134	46%	494	JF751111.1	1565 bp
5	A/swine/Shandong/8/2003(H9N2) segment 5 nucleocapsid protein (NP) gene, partial cds	Influenza A virus	95.57%	1,00E-134	46%	494	GQ477297.1	1476 bp

Tabla S6. Resultados Blastn Secuencia Desconocida 6

Nº	Descripción	Tipo de virus	% Identidad	E-value	Cobertura (%)	Score	Accession	Longitud
1	Ebola virus/M.fascicularis-w/GAB/2011/180mp-CAK1013D3, genoma completo	Ebola virus	37%	8,00E-164	96.69%	18871	KY758965.1	592 bp
2	Ebola virus/M.fascicularis-w/GAB/2011/100mp-CA4505D5, genoma completo	Ebola virus	37%	8,00E-164	96.69%	18871	KY758963.1	592 bp
3	Ebola virus/cynomolgus macaque-w/GAB/2011/10flu-CB145D5, genoma completo	Zaire ebolavirus	37%	8,00E-164	96.69%	18871	KY471095.1	592 bp
4	Ebola virus/M.fascicularis-w/GAB/2011/180mp-CB0201D12, genoma completo	Ebola virus	37%	8,00E-164	96.69%	18871	KY758969.1	592 bp
5	Ebola virus/M.fascicularis-w/GAB/2011/150mp-CA989D5, genoma completo	Ebola virus	37%	8,00E-164	96.69%	18871	KY758948.1	592 bp

Tabla S7. Resultados Blastn Secuencia Desconocida 7

Nº	Descripción	Especie	Score	Longitud	Query Cover	E-value	% Identidad	Accession
1	Dengue virus type 2 strain New Guinea C/PUO-218 hybrid polyprotein gene, complete cds	Dengue virus type 2	448	10724 bp	44%	1,00E-120	91.99%	AF038402.1
2	Dengue virus type 2 isolate DENV2/Senegal/Unknown/Yale-DH083 polyprotein (POLY) gene, complete cds	Dengue virus type 2	694	10512 bp	44%	1,00E-120	91.99%	OR389326.1
3	Dengue virus 2 New Guinea C derivative strain, complete genome	Dengue virus type 2	694	10632 bp	44%	1,00E-120	91.99%	FJ390389.1
4	Dengue virus type 2 strain 43, complete genome	Dengue virus type 2	694	10723 bp	44%	1,00E-120	91.99%	AF204178.1
5	Dengue virus 2 isolate Cuba A35, 1981 polyprotein gene, complete cds	Dengue virus type 2	694	10176 bp	44%	1,00E-120	91.99%	KF704358.1

Tabla S8. Resultados Blastn Secuencia Desconocida 8

Nº	Descripción	Especie	Score	Max Score	% Identidad	E-value	Query Cover	Longitud	Accession
1	A/swine/Shandong/203/HS N1 nucleoprotein gene, complete cds	Influenza A virus	479	479	37%	3,00E-130	96.01%	1565 bp	AY648428.1
2	A/environment/Qinghai/1208/HSN1 segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	479	479	37%	3,00E-130	96.01%	1520 bp	FJ455823.1
3	A/plateau pika/Qinghai/04/2007/HSN1 segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	473	473	37%	1,00E-128	95.68%	1565 bp	FJ390062.2
4	A/chicken/Shandong/LY-12008/H9N2 segment 5 nucleocapsid protein (NP) gene, complete cds	Influenza A virus	473	473	37%	1,00E-128	95.68%	1565 bp	JF755111.1
5	A/swine/Shandong/8/2003/H9N2 segment 5 nucleocapsid protein (NP) gene, partial cds	Influenza A virus	473	473	37%	1,00E-128	95.68%	1476 bp	GQ477297.1

Tabla S9. Resultados Blastn Secuencia Desconocida 9

Nº	Descripción de la secuencia	Especie	Score	Max Score	% Identidad	E-value	Query Cover
1	Homo sapiens RefSeqGene (probablemente INS-IGF2 readthrough)	Homo sapiens	444	887	43%	2,00E-119	99.19%
2	Homo sapiens secuencia genómica completa (posible región reguladora del cromosoma 11)	Homo sapiens	444	887	43%	2,00E-119	99.19%
3	Homo sapiens secuencia genómica con genes TH, INS, IGF2	Homo sapiens	444	887	43%	2,00E-119	99.19%
4	Homo sapiens región genómica rica en loci funcionales	Homo sapiens	444	887	43%	2,00E-119	99.19%
5	Homo sapiens Tyrosine hydroxylase (TH), RefSeqGene	Homo sapiens	444	887	43%	2,00E-119	99.19%

Tabla S10. Resultados Blastn Secuencia Desconocida 10

Nº	Descripción de la cepa	Especie	Tamaño	Nº de Genes	% Identidad	E-value	Accession	Cobertura
1	Serovar Enteritidis strain CVM N175192, cromosoma, genoma completo	<i>Salmonella enterica</i> subsp. <i>enterica</i>	1068	2053	28%	0	CP028726.1	100.00%
2	Strain 143, cromosoma, genoma completo	<i>Salmonella enterica</i>	1068	2053	28%	0	CP015563.1	100.00%
3	Serovar Enteritidis strain 25, cromosoma, genoma completo	<i>Salmonella enterica</i> subsp. <i>enterica</i>	1068	2053	28%	0	CP138557.1	100.00%
4	Serovar Enteritidis strain CFSAN033543, cromosoma, genoma completo	<i>Salmonella enterica</i> subsp. <i>enterica</i>	1068	2053	28%	0	CP028265.1	100.00%
5	Serovar 4,[5],12:i:- L-4065, ADN, genoma completo	<i>Salmonella enterica</i> subsp. <i>enterica</i>	1068	2053	28%	0	AP023131.1	100.00%

Tabla S11. Resultados Blastn Secuencia Desconocida 11

Nº	Cepa	Especie	Tamaño (bp)	Nº de Genes	% GC	E-value	% Identidad	Accession
1	Strain 143, cromosoma, genoma completo	<i>Salmonella enterica</i>	544	1186	52%	2,00E-149	99.68%	CP015683.1
2	Serovar 4,[5],12:i:- L-4605, ADN, genoma completo	<i>S. enterica</i> subsp. <i>enterica</i>	544	1186	52%	2,00E-149	99.68%	AP023131.1
3	Strain 4A845, cromosoma, genoma completo	<i>S. enterica</i> subsp. <i>enterica</i>	544	1186	52%	2,00E-149	99.68%	CP102198.1
4	Serovar 4,[5],12:i:- strain PNCS014868, cromosoma, genoma completo	<i>S. enterica</i> subsp. <i>enterica</i>	544	1186	52%	2,00E-149	99.68%	CP030893.1
5	Serovar Typhimurium str. 140285 substr. GXS275, cromosoma, genoma completo	<i>S. enterica</i> subsp. <i>enterica</i>	544	1186	52%	2,00E-149	99.68%	CP043399.1

Tabla S12. Resultados Blastn Secuencia Desconocida 12

Nº	Descripción	Especie	Longitud	% Identidad	Accession
1	NIMA related kinase 2 (NEK2), mRNA	<i>Pan troglodytes</i>	695 bp	91.46%	NM_001286514.1
2	PREDICTED: NEK2, transcript variant X1, mRNA	<i>Colobus angolensis palliatus</i>	689 bp	91.29%	XM_011938344.1
3	PREDICTED: NEK2, transcript variant X2, mRNA	<i>Colobus angolensis palliatus</i>	689 bp	91.29%	XM_011938446.1

Tabla S13. Resultados Blastn Secuencia Desconocida 13

Nº	Descripción	Tipo de secuencia	Organismo	% Identidad	E-value	Cobertura (%)	Longitud	Alineamiento	Accession
1	Immunoglobulin heavy constant gamma 1 (G1m marker), mRNA (cDNA clone IMAGE:3059913)	mRNA	<i>Homo sapiens</i>	73%	5,00E-138	100.00%	1640 bp	927 bp	BC009840.1
2	Isolate G3, F5_C19_FU_PS42, Heavy immunoglobulin variable region mRNA, partial cds	mRNA (variable region)	<i>Homo sapiens</i>	73%	2,00E-132	98.90%	1575 bp	907 bp	MW176822.1
3	cDNA DKFZp686F17185 (from clone DKFZp686F17185)	mRNA	<i>Homo sapiens</i>	73%	2,00E-132	98.90%	1739 bp	907 bp	CR393860.1
4	Isolate F2, G2_C19_FU_PS32, Heavy immunoglobulin variable region mRNA, partial cds	mRNA (variable region)	<i>Homo sapiens</i>	73%	2,00E-132	98.90%	1554 bp	907 bp	MW176856.1

Análisis en base al E-value, Porcentaje de Identidad y Cobertura:

A través de múltiples búsquedas BLASTn, se comparó la secuencia incógnita con bases de datos genómicas de diversos organismos, incluyendo virus humanos (Influenza A, Dengue, Ébola), bacterias (*Salmonella enterica*), genes humanos (INS, IGF2, TH, NEK2), e inmunoglobulinas. Los resultados fueron evaluados considerando tres criterios clave: E-value, porcentaje de identidad, y cobertura de la consulta.

Los alineamientos con virus de la Influenza A mostraron E-values extremadamente bajos (hasta $2e-136$), porcentajes de identidad superiores al 95%, y coberturas cercanas al 79%, específicamente con el gen NP (nucleoproteína), una región altamente conservada y funcionalmente esencial en el ciclo viral. En contraste, los matches con virus del Dengue tipo 2 también presentaron E-values bajos ($1e-120$) y identidades del 91.99%, pero con menor cobertura (44%). Los resultados relacionados con el virus del Ébola mostraron E-values de $8e-164$ y cobertura del 96.69%, pero con un porcentaje de identidad más bajo (37%), lo que reduce la certeza de asignación. Las coincidencias con Salmonella enterica y genes humanos presentaron identidades moderadas (28–43%) y valores de E más altos o no reportados, lo que sugiere que no corresponden al patógeno en cuestión.

Considerando todos los parámetros, el patógeno que presenta la mayor evidencia de coincidencia biológica con la secuencia incógnita es el virus de la Influenza A, específicamente en su gen de nucleoproteína (NP). Esta conclusión se basa en la combinación de:

En la tabla de Influenza A, los cinco primeros matches presentaron:

- E-value = 0.0
- % Identidad entre 94.35 y 94.78 %
- Query Cover = 79 %

Estos valores indican coincidencias estadísticamente perfectas (E-value cero), con altísima similitud nucleotídica y un alineamiento que cubre la mayor parte de la secuencia.

Por contraste, los hits de Dengue 2 mostraron identidades algo inferiores (≈ 92 %) y coberturas parciales (44 %), y los del virus del Ébola, aunque con E-value muy bajo ($8e-164$) y alta cobertura (96 %), tuvieron identidades muy bajas (37 %). Las comparaciones con Salmonella y genes humanos arrojaron identidades moderadas (28–43 %) o se basaron en regiones no virales.

La combinación de E-value cero, identidad > 94 % y cobertura alta (79 %) en el gen NP de Influenza A supera claramente el soporte estadístico y biológico de cualquier otro match. Por lo tanto, se concluye que el agente patógeno incógnito es el virus de la Influenza A, y la región analizada corresponde al segmento 5 (nucleoproteína NP), esencial para la encapsulación y ensamblaje.

A su vez, también se realizó un alineamiento múltiple entre las secuencias 4, 5 y 8, las cuales por búsqueda blast muestran ser de Virus de Influenza A. El resultado de este alineamiento mostró un porcentaje de identidad del 97%.

T-COFFEE, Version_11.00 (Version_11.00)
Cedric Notredame
CPU TIME:0 sec.
SCORE=974

*
* BAD AVG GOOD

unknown : 93
unknown_1 : 97
unknown_2 : 96
cons : 97

unknown TAAGGAAACTGCACTCAGCGCCTCCCTGGCCAAAAAGCATGGAGTTTTTAAAAGTATGCTCCTTTCAT
unknown_1 -----
unknown_2 -----
cons

unknown AGATAGATAGATAGATAGTAGATAGATGATAGTGCCCGAAGAGAGAGAGAGAGGGAATGCTGAAATTGA
unknown_1 -----GCTCCTTTCGCCCCGAAGAGAGAGAGA-----A
unknown_2 -----
cons

unknown AGATCTCATCTTTCTGGCACGGTCCGCACTCATTCTGAGAGGATCAATGGCCCATAAGTCCTGCCTGCC
unknown_1 GAGCCTGAGCTTTGTGAGGACGTCC-----TGTGGAACCTGTCCCGGGCCCCC
unknown_2 -----
cons

unknown TGCTTGTTGTGTACGGACTTGCCGTAGCCAGTGGATATGACTTCGAGAGAGAAGGGTACTCTTTCAGCAG
unknown_1 TGCCCTGGGATGGGGAGAAAGTCAGGGGGATAGACAGAGTCAAGGTGGGGGACAGGGCGGGAGTGGGGTC
unknown_2 -----GGGGGAGGCCTTTGAAGCAG
cons * * *

unknown	TCCAATCCTTCCAATCCTAGTCGGAATAGATCCTTTCCGTCTGCTGCAAAACAGCCAGGTCTTCAGTCT
unknown_1	CCCAGGGCTGGGGGCCCTTG-----
unknown_2	TCCCT-----
cons	**
unknown	AATTAGACCAAATGAGAATCCAGCACATAAAAGTCAATTGGTGTGGATGGCATGCCATTCTGCAGCGTT
unknown_1	-----
unknown_2	-----
cons	
unknown	TGAAGATCTGAGAGTATCAAGTTTCATCAGAGGAACAAGAGTGGTCCCAAGAGGACAGCTATCTACCAG
unknown_1	-----
unknown_2	-----
cons	
unknown	AGGAGTTCAGATTGCTTCAAATGAGAACATGGAAACAATGGACTCCAGCACTCTTGAAGTGAAGCAG
unknown_1	-GTAGTTCAGATTGCTTCAAATGAGAACATGGAAACAATGGACTCCAGCACTCTTGAAGTGAAGCAG
unknown_2	--GAGTTCAGATTGCTTCAAATGAGAACATGGAAACAATGGACTCCAGCACTCTTGAAGTGAAGCAG
cons	*****
unknown	ATATTGGGCTATAAGGACCAGGAGTGGAG-----
unknown_1	ATATTGGGCTATAAGGACCAGGAGTGGAGGAAACACCAACCAGCAGAGAGCATCTGCAGGACAGATCAG
unknown_2	ATATTGGGCTATAAGGACCAGGAGTGGAGGAAACACCAACCAGCAGAGAGCATCTGCAGGACAGATCAG
cons	*****
unknown	-----
unknown_1	TGTGCAGCCTACTTTCTCAGTACAGAGAAATCTTCCTTTTGAAGAGCGACCATTATGGCGGCATTAC
unknown_2	TGTGCAGCCTACTTTCTCAGTACAGAGAAATCTTCCTTTTGAAGAGCG-----GCATTAC
cons	

unknown	-----
unknown_1	AGGGAATACGGAGGGCAGAACACAGTCCAATCCTTCTGACATGAGGACTGAAATCATAAGAATGATGGA
unknown_2	AGGGAATACGGAGGGCAGAAC-----TCTGACATGAGGACTGAAATCATAAGAATGATGGA
cons	-----
unknown	-----
unknown_1	AAGTGCCAAACCAGAAGACGTGTCTTTCCAGGGGCGGGG-----
unknown_2	AAGTGCCAAACCAGAAGACGTGTCTTTCCAGGGGCGGGGCCAAAAAGCATGAGAAATCAGATATAGAA
cons	-----
unknown	-----
unknown_1	-----
unknown_2	ATCAGATATAGAAATCAGATATAGAAATCAGATATAGAAATCAGATATTCCTGGCCAAAAAGCATGT
cons	-----
unknown	-----
unknown_1	-----
unknown_2	CCCTGGCCATAGATAGATAGATAGATAGATAGATAGAGCCTGAGCTTTGTGAGGACGTCTGTGGA
cons	-----
unknown	-----
unknown_1	-----
unknown_2	ACCTGTCCCGCCCCCTGCCCTGGGATGGGGAGAAGTCAGGGGGATAGACAGAGTCAAGGTGGGGGACA
cons	-----
unknown	-----
unknown_1	-----CAGTGACCAGAGTGTGAGGAGAGGGGAGCA
unknown_2	GGGCGGGAGTGGGGTCCCCAGGGCTGGGGGCCCTTTGGTGAGTGACCAGAGTGTGAGGAGAGGGGAGCA
unknown	-----
unknown_1	AAGCCCTCTAGCCTCATCCTCATAAAAGGTCTCATCATTTTCCCTCCAGCCTCTTATGCACTGGGGACT
unknown_2	AAGCCCTCTAGCCTCATCCTCATAAAAGGTCTCATCATTTTCCCTCCAGCCTCTTATGCACTGGGGAG-
cons	-----
unknown	-----
unknown_1	TTTGAACGAGGCCTTATAGATAGATAGATAGATAGTAGATAGATAGATGATAGTTGAAGCAGGTGTTTCTGGA
unknown_2	-----ATAGTAAAAAGCATGGTGTTGATTTTCAAGAGAGTGCGGACAGTTTCCTTCGA
cons	-----
unknown	-----
unknown_1	GAAATAATCTGAGAGAGAGCATGGTCTTTTCCCTCA-----
unknown_2	GGCCTTTGAAGCAGGTGTCCAGAATTTTACCTTCTTACAATATAGC
cons	-----

Figura S5. Alineamiento múltiple entre las secuencias 4, 5 y 8, las cuales por búsqueda blast muestran ser de Virus de Influenza A.

c) Una vez determinado el agente patógeno usted deberá bajar el genoma completo del organismo en cuestión y realizar una anotación completa de su genoma o de las porciones disponibles. Se deberán presentar la anotación genómica completa.

Se identificó el patógeno en el apartado 3 letra b), concluyendo que era Influenza A. Considerando esto, se descargó el Genoma desde NCBI.

Se realizó la visualización genómica con sus anotaciones respectivas utilizando Geneious Prime. La Figura 3, muestra la anotación de secuencias genéticas del virus de la Influenza A, específicamente de la cepa A de Puerto Rico, con identificadores de secuencia que van desde EF467818 a EF467824.

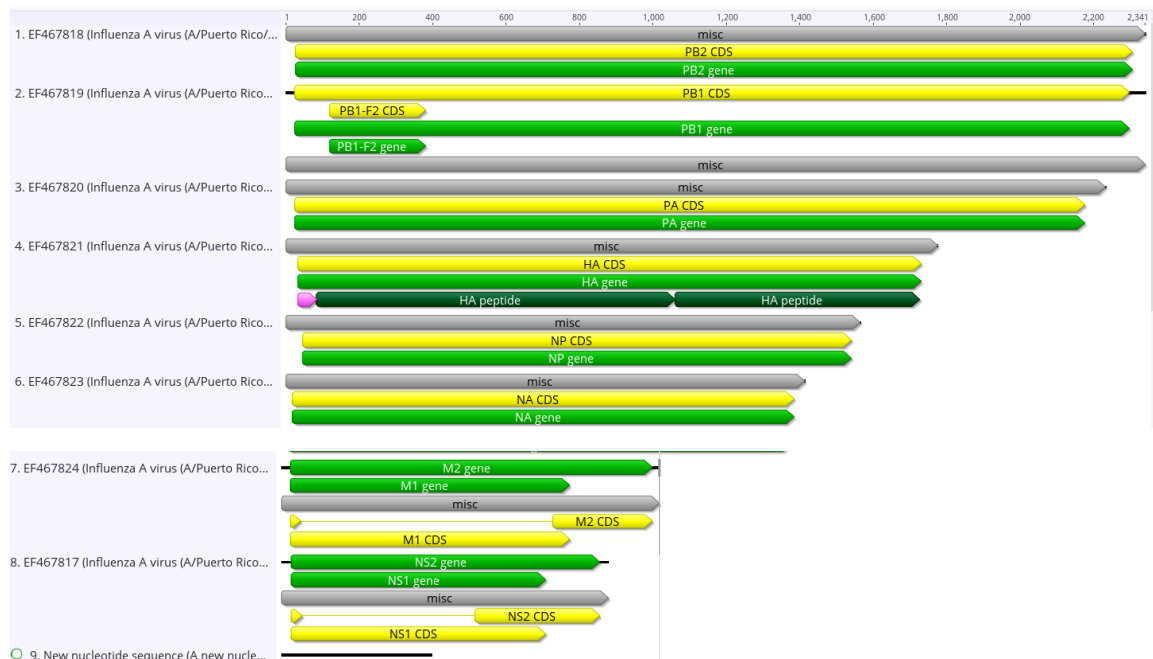


Figura S6: Anotación genómica del Virus Influenza A Puerto Rico.

De acuerdo a esta anotación genómica, podemos resumir que:

Tabla S14. Anotación genómica del Virus Influenza A Puerto Rico.
https://www.ncbi.nlm.nih.gov/datasets/gene/GCF_000865725.1/

Segmento	Proteína codificada	Función principal
PB2	PB2	Subunidad de la ARN polimerasa, implicada en la transcripción viral
PB1	PB1, PB1-F2	Transcripción y replicación del ARN viral; PB1-F2 está asociada a virulencia
PA	PA	Actividad endonucleasa, parte del complejo de replicación
HA	Hemaglutinina (HA)	Unión al receptor celular y entrada al huésped
NP	Nucleoproteína (NP)	Encapsulación del ARN viral
NA	Neuraminidasa (NA)	Liberación de nuevos viriones desde la célula infectada
M	M1, M2	M1: ensamblaje viral; M2: canal iónico para liberación del genoma
NS	NS1, NS2 (NEP)	NS1: inhibición de la respuesta inmune; NS2: exportación nuclear del ARN

4. Análisis de blancos terapéuticos

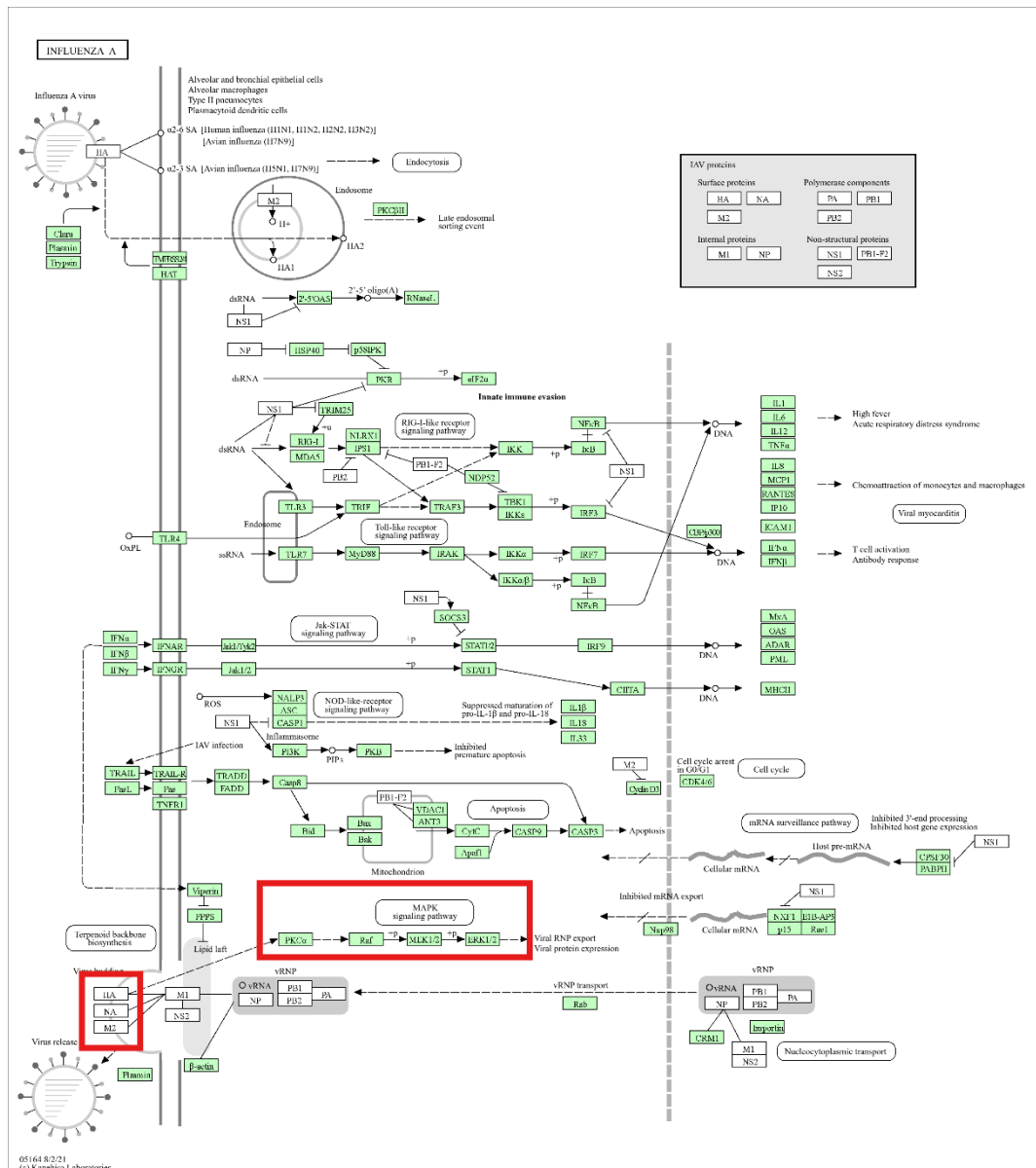


Figura S7. Ruta de señalización asociada a la infección por virus de la Influenza A. Obtenida en: <https://www.kegg.jp/pathway/hsa05164>

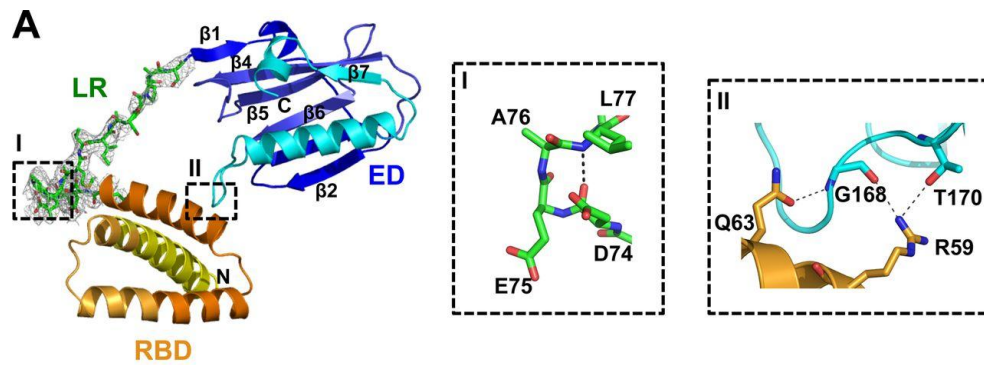


Figura S8. Representación estructural de la proteína NS1 del virus de la Influenza A. Se destacan el dominio de unión al receptor (RBD, naranja), el lazo de reconocimiento (LR, verde) y el dominio efector (ED, azul). Los recuadros I y II amplían las interacciones moleculares clave, incluyendo enlaces de hidrógeno y contactos entre residuos específicos como A76, D74, E75, L77 (caja I) y Q63, G168, T170, R59 (caja II), que son fundamentales para la estabilidad y función de la proteína. (9)

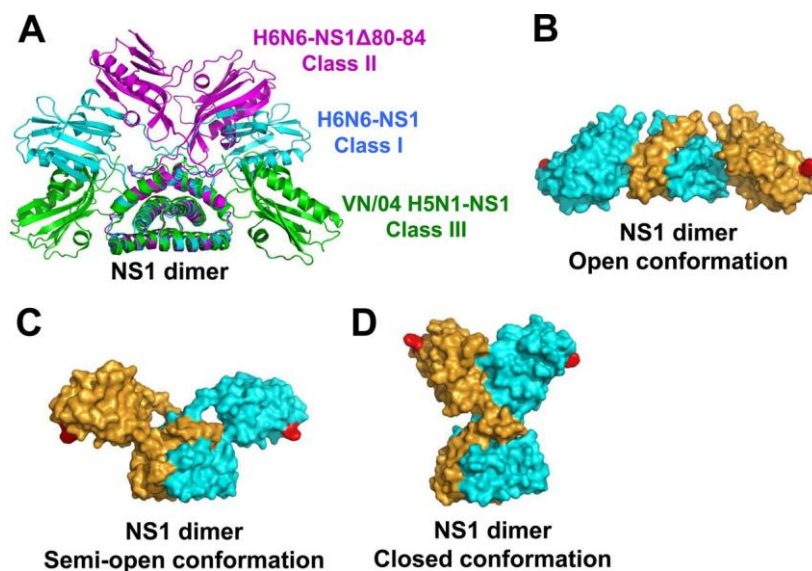


Figura S9. Representaciones estructurales del dímero de la proteína NS1 del virus de la influenza A. (A) Comparación entre las clases estructurales I (azul), II (fucsia) y III (verde) derivadas de distintas cepas virales. (B–D) Modelos de superficie que muestran las conformaciones abierta, semiabierta y cerrada del dímero de NS1, destacando su plasticidad estructural, clave en la evasión inmune y regulación de la respuesta del hospedero. (9)

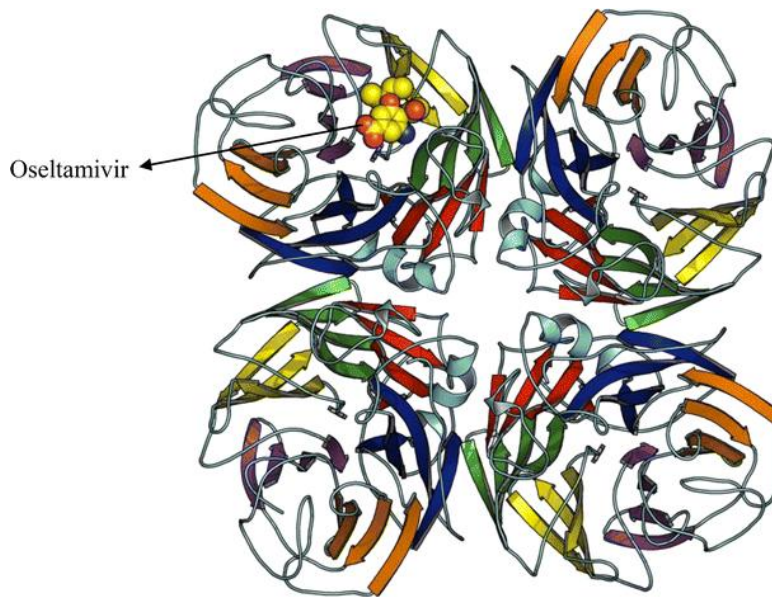


Figura S10. Estructura tridimensional de la proteína neuraminidasa (NA) del virus de la influenza A en forma de tetrámero. Se destacan las hojas β antiparalelas (colores variados) y las hélices (cian). Cada subunidad (~469 aminoácidos) se organiza en cuatro regiones principales: una cola citoplasmática (6 aa), un dominio transmembrana hidrofóbico (aa 7–29), un tallo (~50 aa) y una cabeza globular donde se encuentra el sitio activo enzimático (19 aa). Esta cabeza presenta seis hojas β antiparalelas de cuatro cadenas, dispuestas como aspas de hélice propulsora. El sitio activo, altamente conservado entre cepas, está ubicado en el centro de cada subunidad, formando un bolsillo profundo donde se une oseltamivir y otros inhibidores. (10)

Obtención de estructuras proteicas

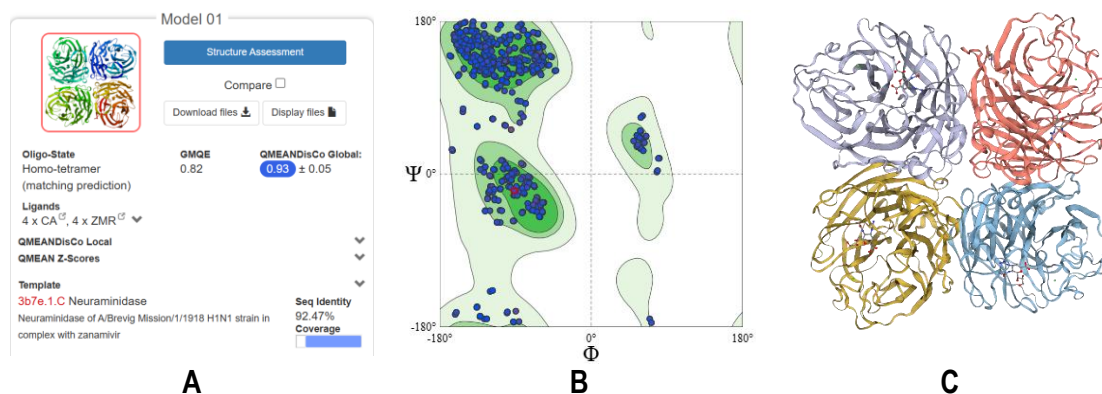


Figura S11. Validación del modelo estructural tridimensional de la neuraminidasa del virus de influenza A (cepa A/Puerto Rico/8/1934(H1N1)). (A) Evaluación de calidad del modelo generado por homología usando Swiss-Model, basado en la plantilla 3b7e.1.C. (B) Diagrama de Ramachandran. (C) Representación tridimensional del modelo homotetramérico generado, con cada monómero coloreado distintamente.

En la Figura S11 se observa el modelamiento de la estructura tridimensional de la neuraminidasa (NA) del virus de influenza A (cepa A/Puerto Rico/8/1934(H1N1)) utilizando la secuencia de referencia NP_040981.1, obtenida de NCBI. El modelo seleccionado presentó una alta calidad estructural con un valor GMQE de 0.82 y un QMEANDisCo Global de 0.93 ± 0.05 . La plantilla

empleada fue la neuraminidasa cristalizada del subtipo H1N1 (PDB ID: 3b7e, cadena C), con una identidad de secuencia del 92.47 %. La estructura generada mostró un estado oligomérico homotetramérico, consistente con la conformación biológica conocida de NA. La validación estructural reveló un MolProbity score de 1.14, un Clash Score bajo (1.06) y 0 % de outliers en el diagrama de Ramachandran, con un 95.17 % de residuos en regiones favorecidas, indicando una buena calidad estereoquímica del modelo. Este modelo fue empleado posteriormente en los estudios de acoplamiento molecular.

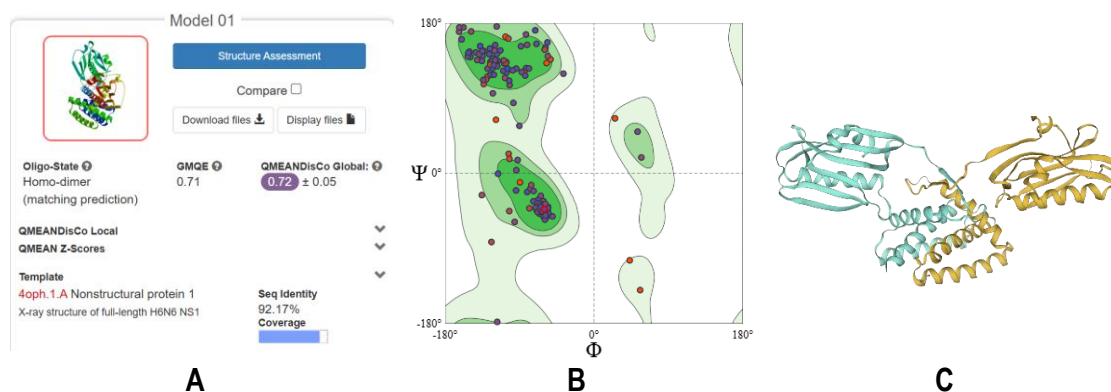


Figura S12. Validación estructural del modelo tridimensional de la proteína no estructural 1 (NS1) de influenza A (Puerto Rico/8/1934 H1N1). (A) Información general del modelo generado por homología en SWISS-MODEL, utilizando como plantilla la proteína NS1 del subtipo H6N6 (PDB ID: 4oph.1.A), con 92.17% de identidad de secuencia. (B) Diagrama de Ramachandran. (C) Visualización de la estructura en forma de homodímero, evidenciando la conformación esperada de la proteína.

El modelo tridimensional de la proteína NS1 del virus de la influenza A se generó mediante homología usando SWISS-MODEL, tomando como plantilla la estructura cristalográfica del NS1 completo del subtipo H6N6 (PDB ID: 4oph.1.A), con un 92.17% de identidad de secuencia y una cobertura elevada. El modelo obtenido presentó un valor GMQE de 0.71 y una puntuación QMEANDisCo Global de 0.72 ± 0.05 , lo que indica una calidad moderadamente confiable. El análisis estructural mostró un dímero como estado oligomérico predominante, consistente con datos experimentales. El análisis del diagrama de Ramachandran indicó que el 95.50% de los residuos se ubicaron en regiones favorecidas, y solo el 0.50% en regiones fuera de conformación permitida, específicamente dos residuos de prolina. El score MolProbity fue 2.01, con un Clash Score de 2.18, lo cual sugiere un modelo adecuado para estudios posteriores. La validación estructural y visualización tridimensional confirman la integridad del modelo (ver Figura S12).

Predicción de Sitios Activos

Tabla S15. Características de los bolsillos identificados en la proteína NS1 mediante la herramienta PockDrug.

Bolsillo	Volumen Hull (Å ³)	Hidrofobicidad (Kyte)	Residuos Polares	Residuos Aromáticos	Átomos Otyr	Nº Residuos	de Prob. Druglikeness	de Desviación Estándar
2	1821.28	0.58	0.44	0.12	0.0	25	0.95	0.00
3	1421.95	0.09	0.50	0.14	0.0	22	0.84	0.01
1	2071.98	-0.65	0.57	0.07	0.0	28	0.43	0.01
0	3598.42	-0.72	0.59	0.06	0.0	34	0.42	0.03
4	1487.39	-0.82	0.56	0.00	0.0	16	0.24	0.05
7	1122.83	-1.62	0.78	0.11	0.0	18	0.06	0.00

En la tabla S15 se muestran el volumen (Vol. Hull), la hidrofobicidad media (Kyte), proporción de residuos polares y aromáticos, número de átomos tipo oxígeno y tirosina (Otyr), número de residuos totales, probabilidad de ser un bolsillo con características tipo fármaco (Druglikeness) y su desviación estándar.

Sin embargo, en el análisis de drugabilidad de la neuraminidasa (ver Tabla S16), se observó que el bolsillo 0, correspondiente al sitio activo donde se unen antivirales como el oseltamivir, mostró una puntuación inferior (0.09) frente al bolsillo 2 (0.74). Esta diferencia puede explicarse por la naturaleza polar del bolsillo catalítico, que reduce su puntuación en herramientas como PockDrug. Sin embargo, este sitio sigue siendo funcionalmente relevante, ya que aloja fármacos clínicamente aprobados. Esto pone de manifiesto que, aunque los algoritmos computacionales son útiles para priorizar sitios de unión, no siempre identifican como más promisorio al sitio activo real, resaltando la importancia de integrar datos estructurales, conocimiento farmacológico previo y validación experimental en la selección de blancos terapéuticos.

Tabla S16. Características de los bolsillos identificados en la proteína NA mediante la herramienta PockDrug.

Bolsillo	Volumen (Å ³)	Hidrofobicidad (Kyte)	Residuos Polares	Residuos Aromáticos	Átomos tipo Otyr	Nº Residuos	de Prob. Farmacológica	Desv. Estándar
0	1629.24	-1.58	0.80	0.12	0.01	25.0	0.09	0.02
1	1705.45	-1.90	0.74	0.04	0.00	27.0	0.03	0.01
2	450.62	0.11	0.53	0.07	0.00	15.0	0.74	0.04

Acoplamiento Molecular y Cribado Virtual

Inicialmente, se descargó una base de datos compuesta por 281 metabolitos de origen natural de la base de datos PeruNPDB (<https://perunpdb.com.pe/>). Posteriormente, se realizó un proceso de filtrado utilizando el software DataWarrior, aplicando criterios relacionados con propiedades farmacológicas y toxicológicas. Los compuestos fueron evaluados en función de su *druglikeness*,

manteniéndose únicamente aquellos con valores iguales o superiores a 0, lo que sugiere una mayor probabilidad de poseer características compatibles con fármacos. Además, se descartaron aquellos compuestos que presentaban alertas de mutagenicidad, tumorigenicidad, efectos en la reproducción, irritabilidad, así como aquellos que contenían patrones PAINS (Pan Assay INterference compoundS), con el fin de asegurar la selección de candidatos confiables y con bajo riesgo de toxicidad o resultados falsos positivos en ensayos biológicos.

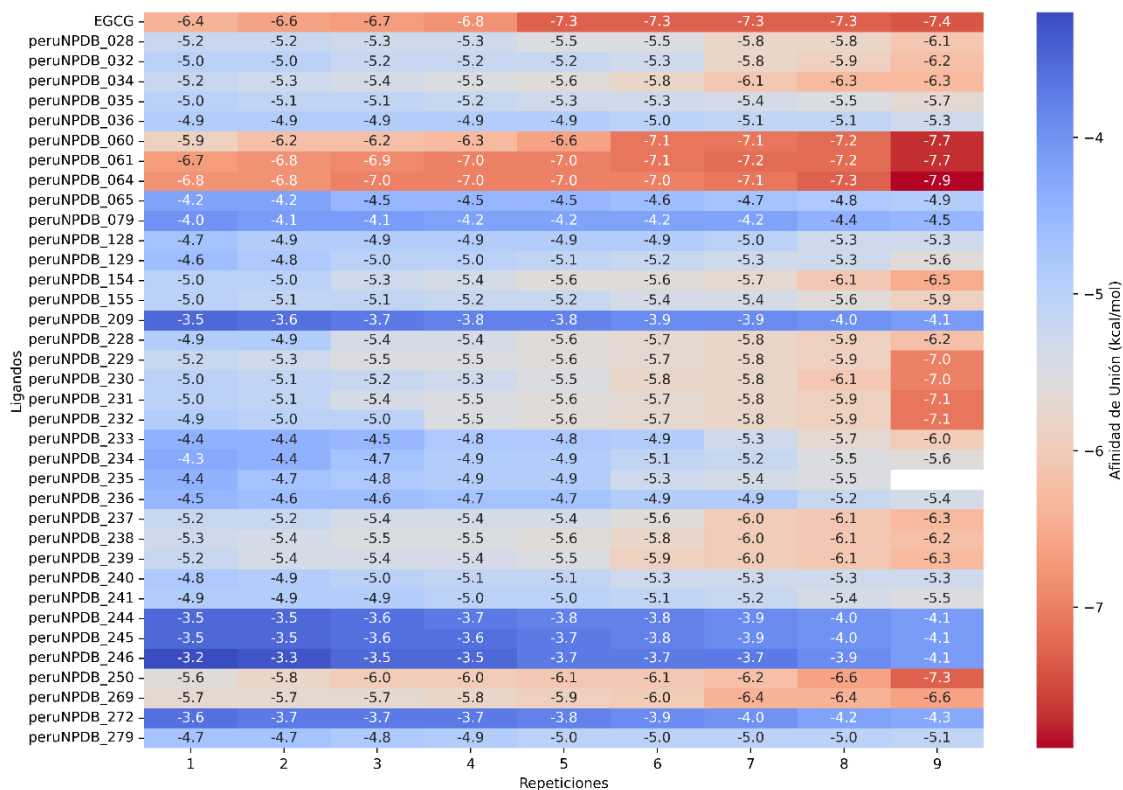


Figura S13. Mapa de calor de las afinidades de unión (en kcal/mol) obtenidas mediante acoplamiento de 37 contra la proteína NS1 del virus de la influenza. Cada fila representa un ligando y cada columna una repetición independiente del docking.

Se realizó un cribado virtual de 36 compuestos naturales de la base de datos peruNPDB contra la proteína no estructural 1 (NS1) del virus de la influenza A. Como control positivo se empleó el compuesto epigallocatequina galato (EGCG),(11) seleccionado a partir de una revisión previa en la que fue identificado como uno de los inhibidores más eficaces de NS1. El análisis de la afinidad de unión mostró que varios compuestos presentaron mejores valores que el control. En particular, peruNPDB_064 alcanzó una afinidad máxima de -7.9 kcal/mol, superando consistentemente al EGCG (-7.4 kcal/mol). Asimismo, peruNPDB_060 y peruNPDB_061 evidenciaron afinidades destacadas (-7.7 y -7.9 kcal/mol en sus mejores repeticiones, respectivamente). Estos resultados sugieren que estos compuestos podrían tener un alto potencial como inhibidores de NS1 y ameritan su evaluación en análisis de dinámica molecular y estudios experimentales posteriores.

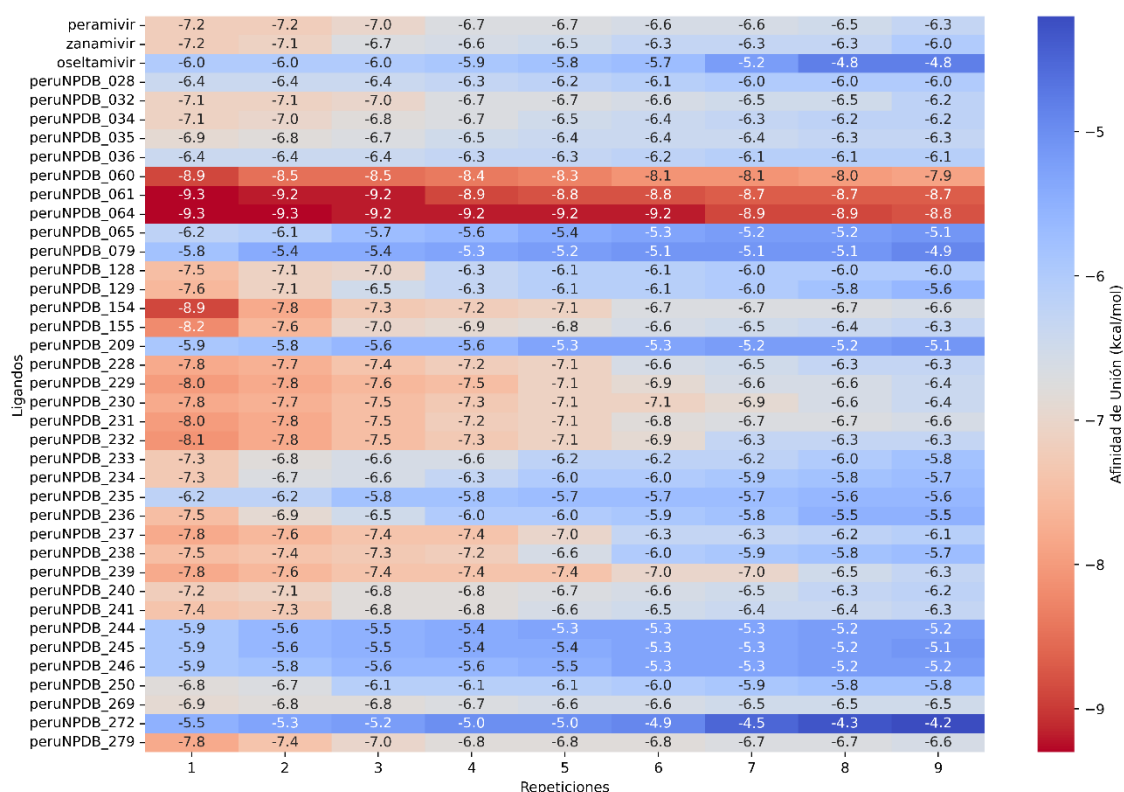


Figura S14. Mapa de calor de las afinidades de unión (en kcal/mol) obtenidas mediante acoplamiento de 39 compuestos contra la proteína NA del virus de la influenza. Cada fila representa un ligando y cada columna una repetición independiente del docking.

Se evaluó la afinidad de unión de 36 compuestos de la base de datos peruNPDB contra la neuraminidasa (NA) del virus de la influenza A, utilizando como controles positivos los antivirales comerciales peramivir, zanamivir y oseltamivir. Los compuestos peruNPDB_064 y peruNPDB_061 destacaron notablemente al mostrar valores de afinidad de unión consistentes en el rango de -9.3 a -8.7 kcal/mol, superando ampliamente a los controles, cuyos valores oscilaron entre -7.2 y -5.2 kcal/mol. Otros compuestos con buena afinidad fueron peruNPDB_060 (-8.9 kcal/mol) y peruNPDB_154 (-8.9 kcal/mol). Estos resultados sugieren que ciertos compuestos naturales peruanos podrían presentar mayor afinidad por la NA viral que los fármacos actualmente aprobados, lo cual refuerza su potencial como posibles inhibidores antivirales para estudios futuros, incluyendo dinámica molecular y validación experimental.

5. Red de interacción génica

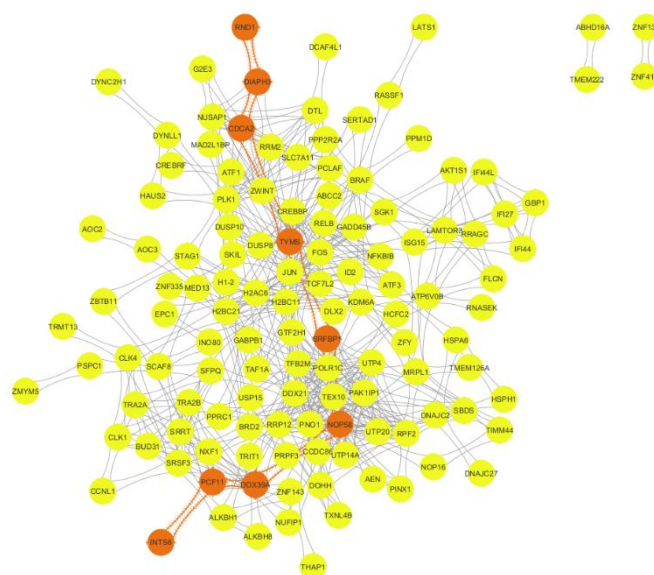


Figura S15. Red de interacción proteína-proteína (PPI) construida a partir de los genes sobreexpresados representando los genes de color naranja que conforman la geodésica.

El análisis de la red de interacción proteína-proteína (PPI) reveló una estructura densa y conectada, con un coeficiente de clustering de 0.391 y un grado promedio de 5.7, lo que sugiere una fuerte modularidad funcional entre los genes sobreexpresados en la infección por Influenza A. La geodésica, conformada por nueve genes, incluyó nodos clave como *PCF11*, *NOP58* y *DDX39A*, los cuales están implicados en la regulación transcripcional, procesamiento de ARN y remodelamiento de cromatina, funciones críticas en la respuesta celular al virus. Además, los genes con alta centralidad de intermediación podrían representar puntos de control estratégicos en la red, ofreciendo potenciales blancos terapéuticos para modular la infección viral desde el contexto del hospedero. Estos resultados concuerdan con estudios previos que destacan la importancia de proteínas nucleares en la replicación viral y la respuesta inmune.

REFERENCIAS BIBLIOGRÁFICAS

1. Thaden JT, Ruffin F, Gjertson D, Hoffmann A, Fowler VG, Yeaman M. 240. Transcriptional signatures differentiate pathogen- and treatment-specific host responses in patients with bacterial bloodstream infections. *Open Forum Infect Dis.* 2023 Nov 27;10(Suppl 2):ofad500.313.
2. Zhang S, Jiang H, Gao B, Yang W, Wang G. Identification of Diagnostic Markers for Breast Cancer Based on Differential Gene Expression and Pathway Network. *Front Cell Dev Biol* [Internet]. 2022 Jan 12 [cited 2025 Jul 22];9. Available from: <https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2021.811585/full>
3. Linggi B, Jairath V, Zou G, Shackelton LM, McGovern DPB, Salas A, et al. Meta-analysis of gene expression disease signatures in colonic biopsy tissue from patients with ulcerative colitis. *Sci Rep.* 2021 Sep 14;11(1):18243.
4. Dexheimer PJ, Pujato M, Roskin KM, Weirauch MT. VExD: a curated resource for human gene expression alterations following viral infection. *G3 (Bethesda).* 2023 Aug 2;13(10):jkad176.

5. Rodriguez-Esteban R, Jiang X. Differential gene expression in disease: a comparison between high-throughput studies and the literature. *BMC Med Genomics*. 2017 Oct 11;10(1):59.
6. Limkar AR, Lack JB, Sek AC, Percopo CM, Druey KM, Rosenberg HF. Differential Expression of Mitosis and Cell Cycle Regulatory Genes during Recovery from an Acute Respiratory Virus Infection. *Pathogens*. 2021 Dec;10(12):1625.
7. Genome guardian p53 and viral infections - Sato - 2013 - Reviews in Medical Virology - Wiley Online Library [Internet]. [cited 2025 Jul 22]. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rmv.1738>
8. Bianco C, Mohr I. Ribosome biogenesis restricts innate immune responses to virus infection and DNA. Sonenberg N, Taniguchi T, Avgousti DC, editors. *eLife*. 2019 Dec 16;8:e49551.
9. Carrillo B, Choi JM, Bornholdt ZA, Sankaran B, Rice AP, Prasad BVV. The Influenza A Virus Protein NS1 Displays Structural Polymorphism. *Journal of Virology*. 2014 Apr 15;88(8):4113–22.
10. Jagadesh A, Salam AAA, Mudgal PP, Arunkumar G. Influenza virus neuraminidase (NA): a target for antivirals and vaccines. *Arch Virol*. 2016 Aug 1;161(8):2087–94.
11. Kim HJ, Jeong MS, Jang SB. Structure and Activities of the NS1 Influenza Protein and Progress in the Development of Small-Molecule Drugs. *International Journal of Molecular Sciences*. 2021 Jan;22(8):4242.