# **Material Suplementario**

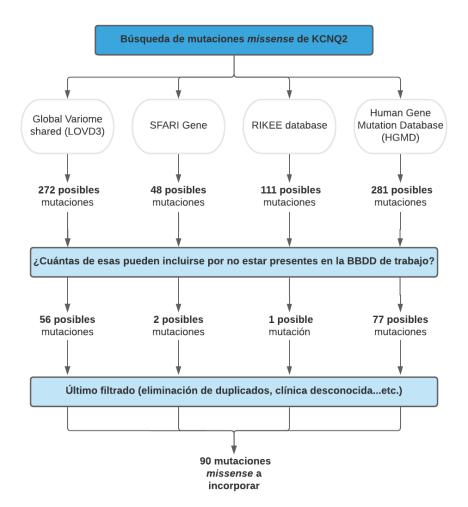
### Ampliación de la base de datos

La base de datos de KCNQ2 fue diseñada a partir del trabajo de Zhang *et al.* (2020) y las mutaciones registradas en ClinVar (<a href="https://www.ncbi.nlm.nih.gov/clinvar/">https://www.ncbi.nlm.nih.gov/clinvar/</a>). Adicionalmente, se realizó una ampliación de la BBDD siguiendo el procedimiento que se muestra a continuación.

En primer lugar, se llevó a cabo una búsqueda bibliográfica exhaustiva para resolver las variantes *missense* de KCNQ2 clasificadas por Zhang *et al.* (2020) y ClinVar como variantes de significado desconocido ("uncertain significance variants" también denominadas como "VUS") o variantes de significado conflictivo ("conflictive interpretation variants).

En segundo lugar, se buscaron nuevas mutaciones *missense* de KCNQ2 en 4 bases de datos adicionales, entre las que se encuentra Global Variome Shared LOVD (<a href="https://databases.lovd.nl/shared/genes">https://databases.lovd.nl/shared/genes</a>), SFARI Gene (<a href="https://gene.sfari.org/">https://gene.sfari.org/</a>), Human Gene Mutation Database o HGMD (<a href="https://www.hgmd.cf.ac.uk/ac/all.php">https://www.hgmd.cf.ac.uk/ac/all.php</a>) y RIKEE database (<a href="https://www.rikee.org/">https://www.rikee.org/</a>). Esta última base de datos es específica para mutaciones de KCNQ2. Posteriormente, se filtraron aquellas mutaciones ya incluidas en la base de datos original. Finalmente, se realizó un último filtrado eliminando posibles duplicados o variantes sin una clínica clara. El procedimiento completo se resume en la **Figura S1**.

**Figura S1.** *Workflow* llevado a cabo para la ampliación de mutaciones de KCNQ2 no presentes en ClinVar empleando otras bases de datos conocidas.



Alba Sáez Matía Trabajo Fin de Máster Material Suplementario

# Diseño de los descriptores de la base de datos

Trece características fueron empleadas para predecir el efecto de las variantes de KCNQ2: el aminoácido previo a la mutación ( $initial\_aa$ ), el aminoácido resultante de la mutación ( $final\_aa$ ), el dominio topológico de Kv7.2 donde tiene lugar la mutación (topological\\_domain), el dominio funcional de Kv7.2 donde ocurre la mutación (functional\\_domain), el valor de conservación evolutiva del residuo mutado ( $residue\_conserv$ ), la estructura secundaria que afecta la mutación ( $secondary\_str$ ); así como los cambios de tamaño ( $d\_size$ ), de hidrofobicidad ( $d\_hf$ ), de volumen ( $d\_vol$ ), de accesibilidad media al solvente ( $d\_msa$ ), de carga ( $d\_charge$ ), de polaridad ( $d\_pol$ ) y de aromaticidad ( $d\_aro$ ) que tienen lugar entre el aminoácido original y el resultante de la mutación.

Todos los descriptores fueron obtenidos de fuentes bibliográficas, a excepción de la conservación evolutiva (*residue\_conserv*) y la estructura secundaria (*secondary\_str*) de los residuos afectados por requerirse de software externo.

## Sobre los dominios topológicos y funcionales afectados...

Los principales dominios topológicos y funcionales de Kv7.2 se obtuvieron directamente de información disponible en Uniprot (<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>) (ID Uniprot: O43526), del PDB 7CR3 que puede ser encontrado en RCSB (<a href="https://www.rcsb.org/">https://www.rcsb.org/</a>) a partir de información actualmente conocida en el laboratorio así como de otras fuentes bibliográficas (**Tablas S1 y S2**).

Dominio topológico	Residuo inicial	Residuo final	Referencia
S1	A92	F112	Uniprot (O43526)
S2	E123	V143	Uniprot (O43526)
<b>S</b> 3	P167	S187	Uniprot (O43526)
S4	A196	W218	Uniprot (O43526)
<b>S</b> 5	L232	L252	Uniprot (O43526)
Poro	A265	P285	Uniprot (O43526)
S6	L292	L312	Uniprot (O43526)
hA	R332	N350	RSCB PDB (7CR3)
hTW	H357	T366	RSCB PDB (7CR3)
hB	G535	L559	RSCB PDB (7CR3)
hC	D563	G594	RSCB PDB (7CR3)
hD	R622	R647	Bernardo-Seisdedos (2015)

**Tabla S1.** Principales dominios topológicos de Kv7.2.

**Tabla S2.** Principales dominios funcionales de Kv7.2.

Dominio topológico	Residuo inicial	Residuo final	Referencia	
voltage_sensor_domain (S1-S4)	A92	W218	Ambrosino et al. (2015)	
selectivity_filter	T277	D282	Uniprot (O43526)	
pore_domain (S5-S6)	L232	L312	Bañales-Belaunde (2019);	
			Yus-Nájera et al. (2002)	
CaM_interaction (hA-hTW-hB)	R332	L559	Ambrosino et al. (2015)	
SID_domain(hC-hD)	D563	R647	Bernardo-Seisdedos (2015)	
unknown_function	Residuos que no se incluyan en ninguno de los anteriores rangos			
dilkilowii_function	residuos que n	o se merayan en ming	guno de los unteriores rangos	

#### Sobre los descriptores de cambio...

• **Descriptores cuantitativos** (*d\_size*, *d\_hf*, *d\_vol*, *d\_msa*). Estos descriptores fueron calculados sustrayendo el valor del aminoácido original del resultante de la mutación. El peso molecular (Da) de los aminoácidos fue obtenido del trabajo realizado por Lide (2007), la hidrofobicidad de Kyte &

Doolittle (1982) y tanto el volumen como la accesibilidad media al solvente de Bogardt *et al.* (1980) (**Tabla S3**).

• **Descriptores cualitativos** (*d\_charge*, *d\_pol*, *d\_aro*). Para determinar el cambio de carga a pH fisiológico se clasificaron los aminoácidos en función de esta y luego se generó un código único para cada uno de los posibles cambios. Ej.: Si el aminoácido inicial es positivo (*pos*) y el resultante de la mutación es negativo (*neg*) el cambio se registra como *pos\_to\_neg*, si el aminoácido inicial es neutro (*neu*) y el resultante de la mutación es positivo (*pos*) el cambio se registra como *neu\_to\_pos*, etc. Se realiza el mismo procedimiento para determinar el cambio de polaridad y el cambio de aromaticidad. Tanto la carga de los aminoácidos como su polaridad y su aromaticidad fueron obtenidas del trabajo de Lodish *et al*. (2012). El pH fisiológico de una célula nerviosa es de 7.4 (Jalalvand *et al*., 2016).

Tabla S3, C	Características	fisico-o	uímicas	de los	aminoácid	os a n	H = 7.4.

Aminoácido	Carga	Polaridad	Aromaticidad	Tamaño (Da)	Volumen	Accesibilidad media	Hidrofobicidad
					estandarizado (%)	al solvente (%)	
(Ser) S	neu	polar	non_aromatic	105.09	18.1	40.5	-0.8
(Thr) T	neu	polar	non_aromatic	119.12	34.0	35.3	-0.7
(Gln) Q	neu	polar	non_aromatic	146.15	51.3	43.6	-3.5
(Asn) N	neu	polar	non_aromatic	132.12	35.4	46.1	-3.5
(Tyr) Y	neu	non_polar	aromatic	181.19	78.5	30.1	-1.3
(Cys) C	neu	polar	non_aromatic	121.16	28.0	7.4	2.5
(Gly) G	neu	polar	non_aromatic	75.07	0.0	54.0	-0.4
(Ala) A	neu	non_polar	non_aromatic	89.09	15.9	37.4	1.8
(Val) V	neu	non_polar	non_aromatic	117.15	47.7	19.6	4.2
(Leu) L	neu	non_polar	non_aromatic	131.17	63.6	10.1	3.8
(Ile) I	neu	non_polar	non_aromatic	131.17	63.6	7.5	4.5
(Met) M	neu	non_polar	non_aromatic	149.21	62.8	3.9	1.9
(Pro) P	neu	non_polar	non_aromatic	115.13	41.0	66.2	-1.6
(Phe) F	neu	non_polar	aromatic	165.19	77.2	5.5	2.8
(Trp) W	neu	non_polar	aromatic	204.23	100.0	13.8	-0.9
(Asp) D	neg	polar	non_aromatic	133.10	31.3	45.0	-3.5
(Glu) E	neg	polar	non_aromatic	147.13	47.2	48.6	-3.5
(Lys) K	pos	polar	non_aromatic	146.19	68.0	54.3	-3.9
(Arg) R	pos	polar	non_aromatic	174.20	70.8	50.1	<b>-4.</b> 5
(His) H	pos	polar	non_aromatic	155.16	49.2	28.1	-3.2

#### Sobre la conservación evolutiva y la estructura secundaria afectadas...

El valor de conservación evolutiva de cada residuo (*residue\_conserv*) se calculó realizando un MSA (*Multiple Sequence Alignment*) con Clustal Omega (<a href="https://www.ebi.ac.uk/Tools/msa/clustalo/">https://www.ebi.ac.uk/Tools/msa/clustalo/</a>). La secuencia de 62 especies disponibles en Uniprot (<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>) fueron introducidas en dicho análisis (**Tabla S4**). Especies con longitud de secuencia similar para el canal Kv7.2 (KCNQ2) fueron alineadas y el valor de conservación evolutiva fue posteriormente calculado empleando el software de Capra & Singh (2007): <a href="https://compbio.cs.princeton.edu/conservation/score.html">https://compbio.cs.princeton.edu/conservation/score.html</a>. Los valores de conservación evolutiva obtenidos de los residuos de Kv7.2 se muestran en la **Tabla S4** y las 62 especies empleadas para el análisis en la **Tabla S5.** 

Tabla S4. Valores de conservación evolutiva de los residuos de KCNQ2 (Uniprot ID: O43526) entre especies.

Residuo	Posición	Conservación	Residuo	Posición	Conservación
M	1	0.61970	V	437	0.79104
V	2	0.65838	S	438	0.81378
Q	3	0.71927	L	439	0.79428
K	4	0.74279	K	440	0.82063
S	5	0.73004	D	441	0.77796
R	6	0.74378	R	442	0.83206
N	7	0.68946	V	443	0.81933

Alba Sáez Matía		Trabajo Fin	de Máster	Material Suplementario		
C	0	0.64001	Г	444	0.04624	
G G	8 9	0.64081 0.59054	F S	444 445	0.84624 0.84304	
V	10	0.54742	S	446	0.82763	
Y	11	0.60965	P	447	0.82243	
P	12	0.69383	R	448	0.79558	
G	13	0.67435	G	449	0.70198	
P	14	0.68418	V	450	0.66265	
S	15	0.55969	A	451	0.67856	
G	16	0.50648	A	452	0.67806	
E	17	0.52534	K	453	0.76470	
K	18	0.50083	G	454	0.74590	
K	19	0.49240	K	455 456	0.80623	
L	20	0.47310	G	456 457	0.76045	
K V	21 22	0.52854 0.51094	S P	457 458	0.80282 0.75991	
v G	23	0.58799	Q	450 459	0.70307	
F	24	0.66750	Q A	460	0.51958	
V	25	0.63548	Q	461	0.51320	
Ğ	26	0.64271	T	462	0.55339	
Ĺ	27	0.70767	V	463	0.63261	
D	28	0.67126	R	464	0.74819	
P	29	0.63120	R	465	0.79294	
G	30	0.65360	S	466	0.81409	
A	31	0.65431	P	467	0.84498	
P	32	0.63033	S	468	0.82421	
D	33	0.59923	A	469	0.79260	
<u>S</u>	34	0.54350	D	470	0.78636	
T	35	0.52223	Q	471	0.77468	
R	36	0.66177	S	472	0.77842	
D G	37 38	0.69874	L E	473 474	0.71105	
A	39	0.73496 0.71499	D D	474 475	0.78308 0.76771	
L L	40	0.70248	S	476	0.81098	
L	41	0.69546	P	477	0.84046	
I	42	0.70573	S	478	0.83284	
A	43	0.65683	K	479	0.84257	
G	44	0.66238	V	480	0.82904	
S	45	0.56026	P	481	0.85236	
E	46	0.52478	K	482	0.83295	
A	47	0.42796	S	483	0.83999	
P	48	0.41791	W	484	0.83860	
K	49	0.56875	S	485	0.82450	
R	50	0.62326	F	486	0.83764	
G	51 52	0.65121 0.69629	G	487 488	0.76978 0.77072	
S I	52 53	0.65510	D R	489	0.78869	
L	54	0.64843	S	490	0.70855	
S	55	0.63184	R	491	0.81382	
K	56	0.65363	A	492	0.75318	
P	57	0.70067	R	493	0.82220	
R	58	0.65339	Q	494	0.85722	
A	59	0.56349	A	495	0.79832	
G	60	0.57381	F	496	0.83207	
G	61	0.50685	R	497	0.82022	
A	62	0.54134	I	498	0.70973	
G	63	0.53546	K	499	0.76549	
A	64	0.49678	G	500	0.77879	
G	65 66	0.60756	A	501	0.76190	
K P	66 67	0.63722 0.68279	A S	502 503	0.80566 0.82992	
P P	68	0.71356	S R	503 504	0.82992	
K	69	0.70467	Q Q	504 505	0.86274	
R	70	0.72897	N N	506	0.78213	
N	71	0.72612	S	507	0.70751	
A	72	0.69610	Ē	508	0.62932	

Alba Sáez Matía		Trabajo Fin de Máster			Material Suplementario		
F	73	0.69416	E	509	0.59824		
Y	73 74	0.72552	A	510	0.67193		
R	7 <b>4</b> 75	0.70251	S	511	0.74747		
K	76	0.67593	L	512	0.78220		
L	70 77	0.69504	P	513	0.80217		
Q	78	0.72178	G	514	0.76074		
N N	79	0.72360	E	515	0.74760		
F	80	0.69293	D	516	0.75322		
L	81	0.61123	I	517	0.63781		
Y	82	0.59896	V	518	0.59993		
N	83	0.43987	Ď	519	0.65224		
V	84	0.48211	D	520	0.71512		
Ĺ	85	0.48271	K	521	0.71581		
E	86	0.57100	S	522	0.76492		
R	87	0.63044	Č	523	0.84590		
P	88	0.74223	P	524	0.76275		
R	89	0.72844	C	525	0.85568		
G	90	0.71578	E	526	0.81577		
W	91	0.76343	F	527	0.81931		
A	92	0.71083	V	528	0.77401		
F	93	0.74692	Ť	529	0.73981		
I	94	0.71086	E	530	0.78351		
Y	95	0.76732	D	531	0.81479		
Н	96	0.77205	L	532	0.78477		
A	97	0.73175	T	533	0.81818		
Y	98	0.79362	P	534	0.83474		
V	99	0.74242	G	535	0.77760		
F	100	0.79735	Ĺ	536	0.75804		
Ĺ	101	0.77863	K	537	0.80999		
L	102	0.76658	V	538	0.76296		
V	103	0.80002	S	539	0.77400		
F	104	0.80824	Ī	540	0.80950		
S	105	0.81681	R	541	0.82550		
C	106	0.85110	A	542	0.80048		
L	107	0.78350	V	543	0.79354		
V	108	0.79292	С	544	0.86626		
L	109	0.77893	V	545	0.77938		
S	110	0.80437	M	546	0.86202		
V	111	0.79193	R	547	0.84002		
F	112	0.82657	F	548	0.83488		
S	113	0.80996	L	549	0.78915		
T	114	0.80304	V	550	0.81625		
I	115	0.80977	S	551	0.80222		
K	116	0.75526	K	552	0.82888		
E	117	0.77053	R	553	0.84257		
Y	118	0.80443	K	554	0.84028		
E	119	0.76221	F	555	0.85607		
K	120	0.72616	K	556	0.83927		
S	121	0.75475	E	557	0.83665		
S	122	0.74631	S	558	0.83903		
E	123	0.72438	L	559	0.81458		
G	124	0.70236	R	560	0.84883		
A	125	0.72888	P	561	0.85411		
L	126	0.72318	Y	562	0.86988		
Y	127	0.77728	D	563	0.85911		
I	128	0.73963	V	564	0.83841		
L	129	0.74417	M	565	0.88011		
E	130	0.80182	D	566	0.84738		
I	131	0.77148	V	567	0.83374		
V	132	0.77773	I	568	0.84894		
T	133	0.80987	E	569	0.84468		
I	134	0.81786	Q	570	0.86378		
V	135	0.81173	Y	571	0.86204		
V	136	0.81600	S	572	0.84569		
F	137	0.84209	A	573	0.82470		

Alba Sáez Matía	<u>Trabajo Fin de Máster</u>			Material Suplementario		
G	138	0.81408	G	574	0.81734	
V	139	0.81812	Н	575	0.87221	
E	140	0.82903	L	576	0.80738	
Y	141	0.85121	D	577	0.83355	
F	142	0.79688	M	578	0.86693	
V	143	0.83165	L	579	0.79913	
R	144	0.84035	S	580	0.78338	
I	145	0.82283	R	581	0.83210	
W	146	0.87457	I	582	0.81029	
A	147	0.79082	K	583	0.82548	
A	148	0.82708	S	584	0.80103	
G	149	0.83054	L	585	0.74974	
C	150	0.87656	Q	586	0.74901	
С	151	0.89056	S	587	0.64406	
С	152	0.89513	R	588	0.66296	
R	153	0.86800	V	589	0.70078	
Y	154	0.88678	D	590	0.78494	
R	155	0.86497	Q	591	0.85222	
G	156	0.82749	Ī	592	0.82961	
W	157	0.89334	V	593	0.82265	
R	158	0.84422	G	594	0.79464	
G	159	0.81661	R	595	0.79581	
R	160	0.84997	G	596	0.75567	
L	161	0.80405	P	597	0.67580	
K	162	0.83310	A	598	0.66703	
F	163	0.84854	I	599	0.70066	
A	164	0.82071	T	600	0.71874	
R	165	0.85153	D	601	0.80121	
K	166	0.85337	K	602	0.78879	
P	167	0.86041	D	603	0.80683	
F	168	0.86208	R	604	0.80612	
C	169	0.88252	T	605	0.63462	
V	170	0.83160	K	606	0.77249	
Ĭ	171	0.84385	G	607	0.69574	
D	172	0.84773	P	608	0.66474	
I	173	0.80309	A	609	0.70432	
M	174	0.85607	E	610	0.71386	
V	175	0.80831	A	611	0.65708	
L L	176	0.79154	E	612	0.77115	
I	177	0.77401	L	613	0.71805	
A	178	0.78984	P	614	0.77089	
S	179	0.80554	E	615	0.78590	
I	180	0.80296	D	616	0.82801	
Å	181	0.75635	P	617	0.84615	
V	182	0.79708	S	618	0.84275	
L L	183	0.77764	M	619	0.88422	
A	184	0.76103	M	620	0.87769	
A	185	0.79684	G	621	0.81022	
G	186	0.79120	R	622	0.83746	
S	187	0.78411	L	623	0.79490	
	188	0.84800	G	624	0.75039	
Q	189	0.81077	K	625	0.73039	
G			V			
N	190	0.84766		626	0.80865	
V	191	0.82801	E	627	0.82266	
F	192	0.84783	K	628	0.80592	
A	193 104	0.81929	Q	629	0.83384	
T	194	0.82264	V	630	0.80001	
S	195	0.82273	L	631	0.70846	
A	196	0.80430	S	632	0.76364	
L	197	0.75503	M	633	0.77189	
R	198	0.82569	E	634	0.73121	
S	199	0.82131	K	635	0.71495	
L	200	0.79274	K	636	0.74626	
	201	0.84353	L	637	0.70963	
R F	201 202	0.84811	D	637 638	0.73671	

Alba Sáez Matía		Trabajo Fin	de Máster	Material Suplementario		
L	203	0.80278	F	639	0.74417	
Q	203	0.85940	r L	640	0.70391	
Q I	205	0.83876	V	641	0.72659	
L	206	0.81167	N	642	0.70343	
R	207	0.85191	I	643	0.70858	
M	208	0.88118	Y	644	0.75633	
I	209	0.84613	M	645	0.70608	
R	210	0.86252	Q	646	0.76508	
M	211	0.88965	R	647	0.74660	
D	212	0.85085	M	648	0.76628	
R	213	0.84835	G	649	0.71980	
R	214	0.84763	Ī	650	0.67970	
G	215	0.82331	P	651	0.71743	
G	216	0.82260	P	652	0.69726	
T	217	0.83901	T	653	0.62658	
W	218	0.87503	Ē	654	0.71308	
K	219	0.83142	T	655	0.71047	
L	220	0.80629	E	656	0.68449	
_ L	221	0.80341	Ā	657	0.71780	
G	222	0.79880	Y	658	0.73906	
S	223	0.82258	F	659	0.73216	
V	224	0.81382	G	660	0.68716	
V	225	0.82573	A	661	0.59854	
Y	226	0.86404	K	662	0.60503	
A	227	0.82537	E	663	0.51886	
Н	228	0.87637	P	664	0.55157	
S	229	0.84016	E	665	0.58567	
K	230	0.82729	P	666	0.69463	
E	231	0.82851	A	667	0.73156	
L	232	0.79232	P	668	0.76551	
V	233	0.77404	P	669	0.77228	
T	234	0.83568	Y	670	0.78021	
A	235	0.81594	Н	671	0.78082	
W	236	0.87925	S	672	0.73827	
Y	237	0.87002	P	673	0.73743	
I	238	0.83796	E	674	0.65522	
G	239	0.83092	D	675	0.64413	
F	240	0.84961	S	676	0.63492	
L	241	0.80435	R	677	0.57415	
C	242	0.86024	E	678	0.55321	
L	243	0.80036	H	679	0.50938	
I	244	0.82149	V	680	0.35257	
L	245	0.80565	D	681	0.47152	
A	246	0.80424	R	682	0.56888	
S	247	0.82305	Н	683	0.57840	
F	248	0.84188	G	684	0.60209	
L	249	0.80184	С	685	0.65720	
V	250	0.81736	I	686	0.62997	
Y	251	0.85175	V	687	0.59120	
L	252	0.79964	K	688	0.68721	
A	253	0.80417	I	689	0.60867	
E	254	0.81514	V	690	0.62360	
K	255	0.81637	R	691	0.68417	
G	256	0.75477	S	692	0.69459	
E	257	0.77953	S	693	0.59774	
N	258	0.83332	S	694	0.69098	
D	259	0.77117	S	695	0.68403	
Н	260	0.81164	T	696	0.56848	
F	261	0.84291	G	697	0.66882	
D	262	0.76888	Q	698	0.69566	
T	263	0.83447	K	699	0.65899	
Y	264	0.85251	N	700	0.70111	
Α.	265	0.80531	F	701	0.67598	
A						
A D A	266 267	0.84725 0.83389	S A	702 703	0.63064 0.65117	

Alba Sáez Matía		Trabajo Fin de Máster			Material Suplementario		
L	268	0.81455	P	704	0.63333		
W	269	0.88507	P	705	0.53845		
W	270	0.87828	A	706	0.42804		
G	271	0.81171	A	707	0.33483		
L	272	0.80077	P	708	0.45186		
I	273	0.78549	P	709	0.47314		
T	274	0.78146	V	710	0.50582		
L	275	0.75447	Q	711	0.56289		
T	276	0.77933	C	712	0.71446		
T I	277 278	0.79332 0.78844	P P	713 714	0.71804 0.73266		
G	276 279	0.79177	S	714 715	0.75568		
Y	280	0.81632	T	716	0.74417		
G	281	0.76872	S	717	0.72578		
D	282	0.79944	W	718	0.74612		
K	283	0.78856	Q	719	0.67592		
Y	284	0.77557	P	720	0.56963		
P	285	0.80780	Q	721	0.52495		
Q	286	0.76005	S	722	0.30845		
T	287	0.79635	H	723	0.42621		
W	288	0.83935	P	724	0.43604		
N	289	0.80378	R	725 726	0.54037		
G R	290 291	0.77211 0.79372	Q G	726 727	0.59040 0.58205		
L L	292	0.75034	H	727	0.63708		
L	293	0.74788	G	729	0.68395		
A	294	0.76429	T	730	0.62876		
A	295	0.76361	S	731	0.72936		
T	296	0.78176	P	732	0.72778		
F	297	0.79581	V	733	0.65787		
T	298	0.78553	G	734	0.66628		
L	299	0.75815	D	735	0.67839		
I	300	0.77863	Н	736 737	0.61778		
G	301	0.76378	G S	737	0.60725		
V S	302 303	0.77402 0.77749	S L	738 739	0.66095 0.65984		
F	304	0.79736	V	740	0.66588		
F	305	0.79759	Ř	741	0.71180		
A	306	0.75832	I	742	0.66137		
L	307	0.75125	P	743	0.72372		
P	308	0.79440	P	744	0.76479		
A	309	0.76203	P	745	0.73318		
G	310	0.78401	P	746	0.74910		
I	311	0.81146	A	747	0.67086		
L	312	0.78539	Н	748	0.73760		
G	313	0.80228	E	749 750	0.70835		
S G	314 315	0.82267 0.79829	R S	750 751	0.71329 0.70899		
F	316	0.79629	S L	751 752	0.62458		
A	317	0.80975	S	753	0.62360		
L	318	0.80373	A	754	0.60727		
K	319	0.83670	Y	755	0.62417		
V	320	0.82515	G	756	0.57099		
Q	321	0.86930	G	757	0.61883		
E	322	0.85252	G	758	0.60271		
Q	323	0.88018	N	759	0.59563		
H	324	0.89251	R	760 <b>7</b> 81	0.66625		
R	325	0.86843	A	761	0.58570		
Q	326	0.88753	S	762	0.58460		
K H	327	0.86197	M	763 764	0.53144		
н F	328 329	0.88636 0.86687	E F	764 765	0.57581 0.50947		
r E	330	0.84997	r L	766	0.44810		
K	331	0.85537	R	767	0.43576		

Alba Sáez Matía	Trabajo Fin de Máster				Material Suplementario
D	333	0.85028	E	769	0.44710
R N		0.85654	D D	769 770	0.44719 0.41838
P	334 335	0.85016	T	770 771	0.36963
A	336	0.81318	P	771 772	0.48446
A	337	0.81124	G	772 773	0.39527
G	338	0.80873	C	773 774	0.45784
L	339	0.78368	R	77 <del>4</del> 775	0.46464
I	340	0.80753	P P	775 776	0.48966
Q	341	0.84599	P	770 777	0.53256
S	342	0.75125	E	778	0.57115
A	343	0.81259	G	779	0.54675
W	344	0.87974	N	780	0.56080
R	345	0.83277	L	781	0.57495
F	346	0.81752	R	782	0.62359
Y	347	0.87162	D	783	0.61770
Å	348	0.81789	S	784	0.68221
T	349	0.82835	D	785	0.71302
N	350	0.84759	T	786	0.69935
L	351	0.79809	S	787	0.71355
				788	
S	352 353	0.79218	I		0.72628
R	353	0.83002	S	789	0.72552
T	354	0.79713	I	790 701	0.73600
D	355	0.81798	P	791	0.73034
L	356	0.79340	S	792	0.74211
H	357	0.76207	V	793	0.75895
<u>S</u>	358	0.82396	D	794	0.75029
T	359	0.83093	H	795	0.78207
W	360	0.88364	E	796	0.75249
Q	361	0.82192	E	797	0.74811
Y	362	0.87464	L	798	0.73425
Y	363	0.87464	E	799	0.73364
E	364	0.83577	R	800	0.74535
R	365	0.84564	S	801	0.76741
T	366	0.83517	F	802	0.76956
V	367	0.81897	S	803	0.78319
T	368	0.80378	G	804	0.78461
V	369	0.82805	F	805	0.79780
P	370	0.82623	S	806	0.76669
M	371	0.81816	I	807	0.77024
Y	372	0.77736	S	808	0.75023
S	373	0.59502	Q S	809	0.77169
S	374	0.54782	S	810	0.75459
Q	375	0.52380	K	811	0.70336
T	376	0.40535	E	812	0.72413
Q	377	0.41262	N	813	0.67918
T	378	0.37023	L	814	0.65035
Y	379	0.37022	D	815	0.67015
G	380	0.40128	A	816	0.57812
A	381	0.45914	L	817	0.66230
S	382	0.54948	N	818	0.62224
R	383	0.68661	S	819	0.61281
L	384	0.73815	С	820	0.61274
I	385	0.79097	Y	821	0.65718
P	386	0.84102	A	822	0.53493
P	387	0.85405	A	823	0.47340
L	388	0.81305	V	824	0.39541
N	389	0.84928	Å	825	0.52616
Q	390	0.84681	P	826	0.58713
L L	391	0.78870	C	827	0.68273
E	392	0.78575	A	828	0.67071
L	393	0.78719	K	829	0.72604
L L	393 394	0.78346	V	830	0.72604
L					
D	יחר	ስ ዕንንዕሶ	n		
R N	395 396	0.82280 0.82550	R P	831 832	0.77135 0.79156

Alba Sáez Matía	Trabajo Fin de Máster			-	Material Suplementario		
K	398	0.83378	I	834	0.70398		
S	399	0.80528	Ā	835	0.61937		
K	400	0.82181	E	836	0.58794		
S	401	0.81105	G	837	0.57744		
Ğ	402	0.79364	E	838	0.62442		
Ĺ	403	0.79234	S	839	0.68507		
A	404	0.74404	D	840	0.73743		
F	405	0.81288	T	841	0.71555		
R	406	0.78816	D	842	0.74628		
K	407	0.68472	S	843	0.73611		
D	408	0.66353	D	844	0.70781		
P	409	0.60354	L	845	0.73366		
P	410	0.65380	С	846	0.76451		
P	411	0.67547	T	847	0.68392		
E	412	0.68641	P	848	0.76061		
P	413	0.75001	С	849	0.71251		
S	414	0.69041	G	850	0.67492		
P	415	0.68824	P	851	0.67796		
S	416	0.61419	P	852	0.59805		
K	417	0.41752	P	853	0.52417		
G	418	0.37653	R	854	0.47339		
S	419	0.33857	S	855	0.51862		
P	420	0.31154	A	856	0.53656		
С	421	0.33261	T	857	0.61094		
R	422	0.23592	G	858	0.63367		
G	423	0.23244	E	859	0.62538		
P	424	0.27089	G	860	0.62220		
L	425	0.31812	P	861	0.61726		
С	426	0.39539	F	862	0.62613		
G	427	0.37752	G	863	0.60150		
C	428	0.40771	D	864	0.66308		
С	429	0.40906	V	865	0.59772		
P	430	0.37813	G	866	0.63029		
G	431	0.38120	W	867	0.77927		
R	432	0.41429	A	868	0.59610		
S S	433	0.45711	G	869	0.64025		
S	434	0.53396	P	870	0.63782		
Q	435	0.75031	R	871	0.53425		
K	436	0.76496	K	872	0.45264		

**Tabla S5.** Especies de las que se obtuvieron las secuencias de KCNQ2 para el cálculo de los valores de conservación de sus aminoácidos. Las secuencias de todas especies (62) oscilan entre los 850 y 875 aminoácidos.

Código Uniprot	Especie	Código Uniprot	Especie
A0A6J1YBD7	Acinonyx jubatus	A0A6J1VM80	Notechis scutatus
A0A7N5JUP3	Ailuropoda melanoleuca	A0A7L4BX40	Nyctiprogne leucopyga
A0A2K5EWT1	Aotus nancymaae	A0A6P3FS62	Octodon degus
A0A6P3GU07	Bison bison bison	A0A2U3WMF7	Odobenus rosmarus divergens
G3N2R2	Bos taurus	A0A6J0W124	Odocoileus virginianus texanus
F7CY15	Callithrix jacchus	A0A7L1ERT0	Oenanthe oenanthe
A0A452E7F3	Capra hircus	H0X1I7	Otolemur garnettii
A0A286XK97	Cavia porcellus	A0A6P3YM10	Ovis aries
A0A2K5QX14	Cebus imitator	A0A2R9AU86	Pan paniscus
A0A2K5L1K5	Cercocebus atys	H2QKS4	Pan troglodytes
A0A7L3EB03	Chaetops frenatus	A0A6P9E1I7	Pantherophis guttatus
A0A6J2VKS8	Chanos chanos	A0A2I3NHI2	Papio anubis
A0A6I9KEP6	Chrysochloris asiatica	A0A6J0CPL8	Peromyscus maniculatus bairdii
A0A6J2PDX4	Cottoperca gobio	A0A6P5JXG8	Phascolarctos cinereus
A0A2Y9Q6H4	Delphinapterus leucas	A0A6J2MPK5	Phyllostomus discolor
A0A4W4F1P1	Electrophorus electricus	A0A2Y9T7Y1	Physeter macrocephalus
F6VHK7	Equus caballus	A0A2K6ET95	Propithecus coquereli
A0A1S3WCY9	Erinaceus europaeus	O88943	Rattus norvegicus
M3WAG9	Felis catus	A0A671F4Z0	Rhinolophus ferrumequinum
G3QZ98	Gorilla gorilla gorilla	A0A2K6P629	Rhinopithecus roxellana
A0A7K9DPE3	Hemiprocne comata	A0A7N4NU17	Sarcophilus harrisii

O43526	Homo sapiens	A0A672MMS1	Sinocyclocheilus grahami
A0A2D0RWX1	Ictalurus punctatus	A0A286ZS06	Sus scrofa
A0A287DBX6	Ictidomys tridecemlineatus	A0A6J3Q231	Tursiops truncatus
A0A6J0IPQ1	Lepidothrix coronata	A0A452SKB9	Ursus americanus
A0A340XT99	Lipotes vexillifer	A0A3Q7W3Y1	Ursus arctos horribilis
A0A2K6C5Y1	Macaca nemestrina	A0A6J3BGW9	Vicugna pacos
A0A1U8BXY3	Mesocricetus auratus	A0A4X2LHR0	Vombatus ursinus
F6WQP9	Monodelphis domestica	A0A3Q7U035	Vulpes vulpes
B7ZBV9	Mus musculus	A0A6I8QM30	Xenopus tropicalis
A0A341BVZ4	Neophocaena asiaeorientalis	A0A6J2FFJ4	Zalophus californianus

Para generar el descriptor de estructura secundaria fue necesario combinar la información estructural de KCNQ2 resuelta hasta la fecha (PDB ID: 7CR3) con programas bioinformáticos de predicción de estructura secundaria para completar las zonas estructuralmente desconocidas del canal. Aunque existen diversos programas para tal fin, se decidió trabajar con el metaserver PROTEUS2 (<a href="http://www.proteus2.ca/proteus2/">http://www.proteus2.ca/proteus2/</a>) por permitir recibir como input el fichero PDB y modelar a partir de él mediante homología (<a href="https://www.proteus2.ca/proteus2/">homology modelling</a>). La estructura secundaria de cada residuo de KCNQ2 se muestran el la **Figura S2**.

**Figura S2.** Predicción de la estructura secundaria mediante PROTEUS2. La primera línea hace referencia a la secuencia proteica del canal, la segunda a la estructura secundaria de la que forman parte (C = *coiled coil*, H = *helix*, T = *transmembrane helix* y E = *beta strand*) y la última a la puntuación de la predicción en cada posición, siendo 0 el valor más bajo y 1 el valor más alto.

1	MVQKSRNGGVYPGPSGEKKLKVGFVGLDPGAPDSTRDGALLIAGSEAPKRGSILSKPRAG CCCCCCCCCCCCHHHHHHHHCCECCCCCCCCCCCCCCC	60
61	GAGAGKPPKRNAFYRKLQNFLYNVLERPRGWAFIYHAYVFLLVFSCLVLSVFSTIKEYEK CCCCCCCCCCHHHHHHHHHHHHHHHCCCCHHHTTTTTTTT	120
121	SSEGALYILEIVTIVVFGVEYFVRIWAAGCCCRYRGWRGRLKFARKPFCVIDIMVLIASI HHHHHCTTTTTTTTTTTTTTTTTTTTTTT 666789999999999999999999999999999999999	180
181	AVLAAGSQGNVFATSALRSLRFLQILRMIRMDRRGGTWKLLGSVVYAHSKELVTAWYIGF TTTTTTTCHHHHHHHHHHHCTTTTTTTTTTTTTTTCCHHHHHH	240
241	LCLILASFLVYLAEKGENDHFDTYADALWWGLITLTTIGYGDKYPOTWNGRLLAATFTLI TTTTTTTTTTTTTTTTCCCCCCCCCTTTTTTTTTTT	300
301	GVSFFALPAGILGSGFALKVQEQHRQKHFEKRRNPAAGLIQSAWRFYATNLSRTDLHSTW TTTTTTTTTTCTTTTTTTCHHHHHHHHHHHHHHHHHH	360
361	QYYERTVTVPMYSSQTQTYGASRLIPPLNQLELLRNLKSKSGLAFRKDPPPEPSPSKGSP CCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCC	420
421	CRGPLCGCCPGRSSQKVSLKDRVFSSPRGVAAKGKGSPQAQTVRRSPSADQSLEDSPSKV CCCCCCCCCCCCCEECCCCCCCCCCCCCCCCCCCCCC	480
481	PKSWSFGDRSRARQAFRIKGAASRQNSEEASLPGEDIVDDKSCPCEFVTEDLTPGLKVSI CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	540
541	RAVCVMRFLVSKRKFKESLRPYDVMDVIEQYSAGHLDMLSRIKSLQSRVDQIVGRGPAIT СНИННИНИНИНИНИНИСССССИНИНИНИНИНИНИНИНИН	600
601	DKDRTKGPAEAELPEDPSMMGRLGKVEKQVLSMEKKLDFLVNIYMQRMGIPPTETEAYFG CCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHH	660
661	AKEPEPAPPYHSPEDSREHVDRHGCIVKIVRSSSSTGQKNFSAPPAAPPVQCPPSTSWQP CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	720
721	QSHPRQGHGTSPVGDHGSLVRIPPPPAHERSLSAYGGGNRASMEFLRQEDTPGCRPPEGN CCCCCCCCCCCCCCCEEECCCCHHHHHHHHHHHCCCCCCC	780
781	LRDSDTSISIPSVDHEELERSFSGFSISQSKENLDALNSCYAAVAPCAKVRPYIAEGESD CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	840
841	TDSDLCTPCGPPPRSATGEGPFGDVGWAGPRK 872 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	

En el GitHub (<a href="https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario">https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario</a>) se encuentra el fichero "DDBB\_generation.ipynb" donde se muestra el proceso automatizado para la creación de la base de datos y sus descriptores.

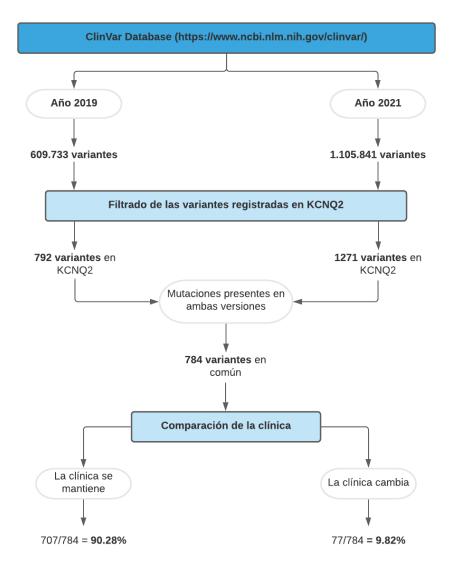
# Estudio de las reclasificaciones de KCNQ2 en ClinVar

El estudio de las reclasificaciones se llevó a cabo descargando los archivos VCF correspondientes a 2019 y 2021, respectivamente:

- https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf GRCh38/archive 2.0/2019/clinvar 20191219.vcf.gz
- https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf\_GRCh38/archive\_2.0/2021/clinvar\_20211204.vcf.gz

Una vez se descargaron los ficheros de ambos años, se filtraron las variantes específicas de KCNQ2. A continuación, se emparejaron las mutaciones de 2019/2021 y se comparó la clínica entre las mutaciones presentes en ambas versiones. El procedimiento completo se muestra en la **Figura S3**.

Figura S3. Workflow llevado a cabo para el estudio de las reclasificaciones de ClinVar de variantes de KCNQ2.



De las 77 variantes que son reclasificadas entre 2019 y 2021, los cambios de etiqueta que suceden son los que se muestran en la **Tabla S6**.

**Tabla S6.** Tabla resumen de las reclasificaciones de KCNQ2 en ClinVar durante los años 2019 y 2021. En la primera columna se muestra la clínica de la mutación en cuestión en el año 2019 y, en la segunda, la clínica en 2021. La última columna recoge el número de veces que ocurre ese cambio entre los años elegidos. La etiqueta "B" hace referencia a "benign", "CI" a "conflicting interpretations", "LB" a "likely benign", "LP" a "likely pathogenic", "P" a "pathogenic" y "VUS" a "variant of uncertain significance".

Clínica 2019	Clínica 2021	N.º veces
В	CI	1
В	LB	2
CI	В	1
CI	LB	5
CI	LP	4
CI	P	3
CI	VUS	1
LB	В	10
LB	CI	4
LP	CI	11
LP	P	3
P	В	2
P	CI	1
P	LP	10
P	VUS	2
VUS	CI	11
VUS	LB	2
VUS	LP	2
VUS	P	2

En el GitHub (<a href="https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario">https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario</a>) se encuentra el fichero "clinic\_reclass\_ClinVar.ipynb" donde se muestra el proceso automatizado para el estudio de las reclasificaciones en ClinVar.

Por último, y siguiendo el mismo procedimiento que para KCNQ2, se realiza el cálculo de las tasas de reclasificación de las variantes anotadas en ClinVar para el resto de miembros de la familia de canales de potasio dependientes de voltaje ( $K_v7.1-K_v7.5$ ). Estos canales se codifican a partir de los genes KCNQ1-KCNQ5, respectivamente. En la **Tabla S7** se muestran los resultados obtenidos.

Tabla S7. Tabla resumen de las tasas de reclasificación estimadas para los miembros de la familia KCNQ.

	KCNQ1		KCNQ2		KCNQ3		KCNQ4		KCNQ5	
	2019	2021	2019	2021	2019	2021	2019	2021	2019	2021
Total genes variants	609733	1105841	609733	1105841	609733	1105841	609733	1105841	609733	1105841
Total KCNQ variants	945*	1373	792	1247	486	821	107	192	16	59
Common variants ('19 - '21)	944		784		486		107		16	
Variants that kept clinic	84.85% (801/944)		90.18% (707/784)		90.95% (442/486)		93.46% (100/107)		81.25% (13/16)	
Variants that changed	15.15% (143/944)		9.82% (	77/784)	9.05% (44/486)		6.54% (7/107)		18.75% (3/16)	

<sup>(\*)</sup> Una variante tuvo que ser eliminada por no tener una clínica definida.

# Modelos de machine learning

Para entrenar cada uno de los modelos se creó un Pipeline() sobre el que se añadió la clase StandardScaler() y uno de los estimadores entrenados. Previamente, se realizó un filtrado de características basado en un *F-score* (ANOVA) mediante la clase SelectKBest(). Los *pipelines* finales para cada uno de los modelos así como el número de características seleccionadas (k) se muestran en la **Tabla S8**.

**Tabla S8.** Número de características seleccionadas y *pipelines* completos de los modelos entrenados. En la tabla se recogen el mejor número de *features* (k) y los parámetros modificados para los modelos de Regresión Lineal con penalización LASSO (L1), *Support Vector Classifier* (SVC) y *Random Forest* (RF). El resto de parámetros que no se muestran mantuvieron su valor por defecto.

Modelo	Nº características (k)	Pipeline
L1	k = 15	pipeline_LR1 = Pipeline( [("scaler", StandardScaler()),
		("logistic_l1", LogisticRegression(solver = "liblinear",
		penalty = "l1",
		fit_intercept = False,
		class_weight = $\{0: 3, 1:1\}$ ,
		C=2.2,
		random_state = 1))] )
SVC	k = 10	<pre>pipeline_SVC = Pipeline( [("scaler", StandardScaler()),</pre>
		("SVC", SVC(kernel = "sigmoid",
		coef0 = 0.3,
		class_weight = $\{0: 3, 1:2\}$ ,
		probability = True,
		random_state = 0))] )
RF	k = "all"	<pre>pipeline_RF = Pipeline( [("scaler", StandardScaler()),</pre>
		("RF", RandomForestClassifier(criterion = "entropy",
		$\max_{depth} = 3$ ,
		min_samples_split = 3,
		n_estimators = 145,
		max_features = "sqrt",
		bootstrap = True,
		max_samples = 15,
		class_weight = $\{0: 9, 1:1\}$ ,
		random state = 12))])

En el GitHub (<a href="https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario">https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario</a>) se encuentra el fichero "ML\_models.<a href="https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario">https://github.com/albasaezmat/TFM\_UAM/tree/main/Material\_suplementario</a>) se encuentra el fichero "ML\_models."

\*\*The properties of the properties of the

#### Predicción del efecto de las variantes

Treinta y ocho variantes de interpretación conflictiva de KCNQ2 anotadas en ClinVar fueron seleccionadas para poner a prueba la precisión de los modelos entrenados. Los resultados se muestran en la **Tabla S9.** 

**Tabla S9.** Predicciones de las 38 mutaciones conflictivas a partir de los modelos diseñados. La primera columna "*Mutations*" hace referencia a la mutación en cuestión. Las siguientes tres columnas ("LR1", "SVC" y "RF") son las predicciones realizadas por los tres modelos entrenados. Los modelos atribuyen la etiqueta 0 (benigna) o 1 (patológica) en función de las probabilidades calculadas para cada clase. Siempre que la probabilidad es  $\geq 0.5$  para la clase benigna, el algoritmo hará la predicción como "clase 0". De manera similar, siempre que la probabilidad sea  $\geq 0.5$  para la clase patológica, el algoritmo hará la predicción como "clase 1. Las últimas tres columnas hacen referencia a las probabilidades calculadas para cada etiqueta por cada uno de los modelos entrenados.

				prob_LR1	prob_SVC	prob_RF
Mutations	LR1	SVC	RF	[class0,class1]	[class0,class1]	[class0,class1]
				[0.99530601,	[0.3785825,	[0.56673998,
P777R	0	0	0	0.00469399]	0.6214175]	0.43326002]
				[0.99653039,	[0.3785825,	[0.67720888,
P777S	0	0	0	0.00346961]	0.6214175]	0.32279112]

Alba Sáez Matía			Trabajo Fin de	Material Suplementario		
				[0.99892292,	[0.14351943,	[0.54208216,
A707T	0	1	0	0.00107708]	0.85648057]	0.45791784]
T0001				[0.52318316,	[0.98931449,	[0.43411991,
T696M	0	0	1	0.47681684]	0.01068551]	0.56588009]
R691H	1	1	0	[0.21712671, 0.78287329]	[0.22780045, 0.77219955]	[0.52121388, 0.47878612]
103111	1	_	Ü	[0.28724608,	[0.39393093,	[0.49933224,
P638L	1	0	1	0.71275392]	0.60606907]	0.50066776]
D. 6007.17				[0.44650922,	[0.23570389,	[0.29378057,
R622W	1	1	1	0.55349078] [0.99779014,	0.76429611] [0.65678528.	0.70621943] [0.69669293,
T605S	0	0	0	0.00220986]	0.34321472]	0.30330707]
	-	-		[0.00627951,	[0.06588571,	[0.25208373,
R563P	1	1	1	0.99372049]	0.93411429]	0.74791627]
DEEDVA	1	1	1	[0.17329038,	[0.22172199, 0.77827801]	[0.30960603, 0.69039397]
R553W	1	1	1	0.82670962] [0.36062588,	[0.34280832,	[0.57424842,
V543M	1	1	0	0.63937412]	0.65719168]	0.42575158]
				[0.11963126,	[0.17496829,	[0.50661173,
E530K	1	1	0	0.88036874]	0.82503171]	0.49338827]
E515D	0	0	0	[0.87606723, 0.12393277]	[0.5, 0.5]	[0.65146637, 0.34853363]
EJIJD	U	U	U	[9.99998380e-01.	[0.57654608,	[0.6640806,
P424S	0	0	0	1.62046151e-06]	0.42345392]	0.3359194]
				[0.99495521,	[0.75653673,	[0.53358947,
S374L	0	0	0	0.00504479]	0.24346327]	0.46641053]
T359A	1	1	1	[0.01201439, 0.98798561]	[0.09506401, 0.90493599]	[0.30691391, 0.69308609]
100311	1	1	1	[0.0974694,	[0.16554593,	[0.50575105,
L356P	1	1	0	0.9025306]	0.83445407]	0.49424895]
P.252.C	0			[0.51028539,	[0.29585997,	[0.36301281,
R353C	0	1	1	0.48971461] [0.16345402,	0.70414003] [0.27783501,	0.63698719] [0.45397837,
H328N	1	1	1	0.83654598]	0.72216499]	0.54602163]
	_	_	_	[0.68801449,	[0.5,	[0.6537805,
G310S	0	0	0	0.31198551]	0.5]	0.3462195]
G301V	0	0	1	[0.70561969, 0.29438031]	[0.49070817, 0.50929183]	[0.42145574,
G301 V	U	U	1	[0.77916056,	[0.57639702,	0.57854426] [0.66929774,
G301S	0	0	0	0.22083944]	0.42360298]	0.33070226]
				[0.13264287,	[0.29463278,	[0.50615374,
D282H	1	1	0	0.86735713]	0.70536722]	0.49384626]
G281R	1	0	1	[0.23790724, 0.76209276]	[0.39954247, 0.60045753]	[0.45612307, 0.54387693]
G2011t	1	Ü	1	[0.35954243,	[0.28771076,	[0.36514075,
F261S	1	1	1	0.64045757]	0.71228924]	0.63485925]
Daroni	0	0	0	[0.68689575,	[0.55040633,	[0.55051588,
D259N	0	0	0	0.31310425] [0.12848652,	0.44959367] [0.25453049,	0.44948412] [0.35908361,
G239V	1	1	1	0.87151348]	0.74546951]	0.64091639]
				[0.31088601,	[0.24072525,	[0.35184739,
T234I	1	1	1	0.68911399]	0.75927475]	0.64815261]
T217I	1	1	1	[0.02626483, 0.97373517]	[0.23139466, 0.76860534]	[0.33269425, 0.66730575]
121/1	1	_	1	[0.08273148,	[0.28567907,	[0.28626122,
R201C	1	1	1	0.91726852]	0.71432093]	0.71373878]
				[0.24943719,	[0.27953164,	[0.311988,
S195F	1	1	1	0.75056281] [0.82044947,	0.72046836] [0.33111732,	0.688012] [0.4588543,
A193V	0	1	1	0.17955053]	0.66888268]	0.5411457]
11100 (	Ü	-	-	[0.97567406,	[0.40129456,	[0.63552224,
V182M	0	0	0	0.02432594]	0.59870544]	0.36447776]
D144547	0	1	1	[1.00000000e+00,	[0.23988083,	[0.28825358,
R144W	0	1	1	1.06172767e-14] [0.94844004,	0.76011917] [0.47060159,	0.71174642] [0.69905012,
V132M	0	0	0	0.05155996]	0.52939841]	0.30094988]
				[0.30896159,	[0.32141595,	[0.32523173,
S113F	1	1	1	0.69103841]	0.67858405]	0.67476827]
A64S	0	1	0	[0.99848584, 0.00151416]	[0.35445448, 0.64554552]	[0.73075382, 0.26924618]
110-13	Ū	1	Ū	[0.92201648,	[0.51932628,	[0.70605748,
A43V	0	0	0	0.07798352]	0.48067372]	0.29394252]
						<del></del> _

En el GitHub (<a href="https://github.com/albasaezmat/TFM">https://github.com/albasaezmat/TFM</a> UAM/tree/main/Material suplementario) se encuentra el fichero "ML\_models.<a href="https://github.com/albasaezmat/TFM">https://github.com/albasaezmat/TFM</a> UAM/tree/main/tree/main/tree/main/tree/main/tree/main/tree/main/

### Cálculo de $\Delta\Delta G$ mediante Rosetta

De las 38 mutaciones conflictivas que se predicen, solo 22 se encuentran en zonas estructuralmente conocidas. Como consecuencia, el cálculo de la inestabilidad o estabilidad proteica que genera una mutación a partir de la variación del cambio en la energía libre de Gibbs ( $\Delta\Delta G$ ) solo puede ser calculada a partir de ellas. Para calcularlo, se sigue el procedimiento detallado por Razquin-Lizarraga (2021).

Debido a los cuatro GAPs que presenta la estructura del canal  $K_v$ 7.2 es necesario renumerar las posiciones que definen las mutaciones en base a ellos (**Tablas S10 y S11**). Esto se debe a que el PDB del canal empleado (PDB: 7CR3) no tiene en cuenta los GAPs en su numeración.

**Tabla S10**. GAPs que presenta la estructura de Kv7.2.

Gap	Residues
1	1 - 69
2	185 -194
3	368 - 534
4	596 - 872

**Tabla S11.** Mutaciones de Kv7.2 con estructura conocida y su renumeración para el cálculo de  $\Delta \Delta G$ . La primera columna hace referencia a la mutación con su nomenclatura estándar. Las dos siguientes columnas son las que necesita Rosetta para hacer el cálculo. Por un lado, "final\_aa" hace referencia al aminoácido que resulta de la mutación. Por el otro lado, "pdb\_position" es la posición real de esa mutación en el PDB sin tener en cuenta los GAPs estructurales. Por último, la columna " $\Delta \Delta G$ " recoge el valor calculado mediante Rosetta.

Mutation	final_aa	pdb_position	ΔΔG
R553W	W	307	5,24
V543M	M	297	3,48
T359A	A	280	0,23
L356P	P	277	63,9
R353C	С	274	4,21
H328N	N	249	-3,95
G310S	S	231	-3,1
G301V	V	222	59,34
G301S	S	222	62,82
D282H	Н	203	3,67
G281R	R	202	517,87
F261S	S	182	53,79
D259N	N	180	0,82
G239V	V	160	31,13
T234I	I	155	-1,81
T217I	I	138	-8,56
R201C	С	122	2,69
S195F	F	116	10,59
V182M	M	113	-7,06
R144W	W	75	38,58
V132M	M	63	8,75
S113F	F	44	-5,88

Alba Sáez Matía Trabajo Fin de Máster Material Suplementario

# Bibliografía

- Ambrosino, P., Alaimo, A., Bartollino, S., Manocchio, L., de Maria, M., Mosca, I., Gomis-Perez, C., Alberdi, A., Scambia, G., Lesca, G., Villarroel, A., Taglialatela, M., & Soldovieri, M. V. (2015). Epilepsy-causing mutations in Kv7.2 C-terminus affect binding and functional modulation by calmodulin. Biochimica et Biophysica Acta (BBA) Molecular Basis of Disease, 1852(9), 1856–1866. <a href="https://doi.org/10.1016/j.bbadis.2015.06.012">https://doi.org/10.1016/j.bbadis.2015.06.012</a>
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., . . . Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871–876. <a href="https://doi.org/10.1126/science.abj8754">https://doi.org/10.1126/science.abj8754</a>
- Bañales-Belaunde, I. (2019). *Regulación del canal de potasio Kv7.2 por la calmodulina*. *Estudio del conector de las hélices A y B en el dominio citoplasmático C-terminal*. [Trabajo de Fin de Grado no publicado]. Instituto Biofísica, CSIC-UPV/EHU, Universidad del País Vasco, Leioa, Spain.
- Bernardo-Seisdedos, G. (2015). *Structural characterization of helices A and B of Kv7.2 channel bound to Calmodulin. The calcium effect.* [Tesis de doctorado no publicada]. Instituto Biofísica, CSIC-UPV/EHU, Universidad del País Vasco, Leioa, Spain.
- Bogardt, R. A., Jones, B. N., Dwulet, F. E., Garner, W. H., Lehman, L. D., & Gurd, F. R. N. (1980). Evolution of the amino acid substitution in the mammalian myoglobin gene. Journal of Molecular Evolution, 15(3), 197–218. https://doi.org/10.1007/bf01732948
- Capra, J. A., & Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics*, 23(15), 1875–1882. <a href="https://doi.org/10.1093/bioinformatics/btm270">https://doi.org/10.1093/bioinformatics/btm270</a>
- Illergård, K., Ardell, D. H., & Elofsson, A. (2009). Structure is three to ten times more conserved than sequence-A study of structural response in protein cores. Proteins: Structure, Function, and Bioinformatics, 77(3), 499–508. <a href="https://doi.org/10.1002/prot.22458">https://doi.org/10.1002/prot.22458</a>
- Jalalvand, E., Robertson, B., Tostivint, H., Wallén, P., & Grillner, S. (2016). The Spinal Cord Has an Intrinsic System for the Control of pH. Current Biology, 26(10), 1346–1351. https://doi.org/10.1016/j.cub.2016.03.048
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., ŽíDek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. <a href="https://doi.org/10.1038/s41586-021-03819-2">https://doi.org/10.1038/s41586-021-03819-2</a>
- Kyte J, Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 1982;157:105–132.
- Lide, D. R. (2007). CRC Handbook of Chemistry and Physics, 88th Edition. Taylor & Francis. Página 1180.
- Lodish, University Harvey, Berk, University Arnold, Kaiser, University Chris A, Krieger, University Monty, Bretscher, University Anthony, Ploegh, University Hidde, Scott, M. P., & Amon, A. (2012). Molecular Cell Biology (7th ed.). W. H. Freeman.
- McInnes, G., Sharo, A. G., Koleske, M. L., Brown, J. E., Norstad, M., Adhikari, A. N., Wang, S., Brenner, S. E., Halpern, J., Koenig, B. A., Magnus, D. C., Gallagher, R. C., Giacomini, K. M., & Altman, R. B. (2021). Opportunities and challenges for the computational interpretation of rare variation in clinically important genes. The American Journal of Human Genetics, 108(4), 535–548. <a href="https://doi.org/10.1016/j.ajhg.2021.03.003">https://doi.org/10.1016/j.ajhg.2021.03.003</a>
- Razquin-Lizarraga, A. (2021). Computational analysis of 12 mutations in the potassium channel KCNQ2 [Trabajo de Fin de Grado no publicado]. Facultad de Ciencia y Tecnología UPV/EHU, Universidad del País Vasco, Leioa, Spain.
- Yus-Nájera, E., Santana-Castro, I., & Villarroel, A. (2002). The Identification and Characterization of a Noncontinuous Calmodulin-binding Site in Noninactivating Voltage-dependent KCNQ Potassium Channels. Journal of Biological Chemistry, 277(32), 28545–28553. <a href="https://doi.org/10.1074/jbc.m204130200">https://doi.org/10.1074/jbc.m204130200</a>
- Zhang, J., Kim, E. C., Chen, C., Procko, E., Pant, S., Lam, K., Patel, J., Choi, R., Hong, M., Joshi, D., Bolton, E., Tajkhorshid, E., & Chung, H. J. (2020). Identifying mutation hotspots reveals pathogenetic mechanisms of KCNQ2 epileptic encephalopathy. *Scientific Reports*, *10*(1). <a href="https://doi.org/10.1038/s41598-020-61697-6">https://doi.org/10.1038/s41598-020-61697-6</a>