VCF data

Position	Ref	Alt
Chr 1: 1	A	-
Chr 1: 2	T	-
Chr 1: 3	G	-
Chr 1: 4	C	-
…		
Chr 1: 11	C	A
Chr 1: 21	T	G
…		
Chr 22:51'324'926	A	T

Compared to reference, no change

Compared to reference, C>A

For the whole genome there are approximately 6,000,000,000 base pairs

After variant effect database

Position	Ref	Alt	Gene	Consequence
Chr 1: 1	A	-	RAG1	None
Chr 1: 2	T	-	RAG1	None
Chr 1: 3	G	-	RAG1	None
Chr 1: 4	C	-	RAG1	None
…				
Chr 1:11	C	A	RAG1	missense
…				
Chr 1: 21	T	G	RAG1	stop

Variant effect checks a database to see if variants occur in gene regions. It may say how that variant will affect the encoded protein.

Most variants have little effect – some can cause protein alterations or stop protein production.

Many databases exist for previously observed variant consequence.

After variant effect database + additional annotation

Position	Ref	Alt	Gene	Consequence	Domain	Clinical_significance
Chr 1: 1	A	-	RAG1	None	Unknown	Unknown
Chr 1: 2	T	-	RAG1	None	Unknown	Unknown
Chr 1: 3	G	-	RAG1	None	Catalytic	Benign
Chr 1: 4	C	-	RAG1	None	Binding	Pathogenic
…						
Chr 1:11	C	A	RAG1	missense	Unknown	Benign
…						
Chr 1: 21	T	G	RAG1	stop	Binding	Pathogenic

Typical human genome

Chromo-some	Length	Base pairs	Varia-tions	Protein-coding genes
1	8.5 cm	248,387,328	12,151,146	2058
2	8.3 cm	242,696,752	12,945,965	1309
3	6.7 cm	201,105,948	10,638,715	1078
4	6.5 cm	193,574,945	10,165,685	752
5	6.2 cm	182,045,439	9,519,995	876
6	5.8 cm	172,126,628	9,130,476	1048
7	5.4 cm	160,567,428	8,613,298	989
8	5.0 cm	146,259,331	8,221,520	677
9	4.8 cm	150,617,247	6,590,811	786
10	4.6 cm	134,758,134	7,223,944	733
11	4.6 cm	135,127,769	7,535,370	1298
12	4.5 cm	133,324,548	7,228,129	1034
13	3.9 cm	113,566,686	5,082,574	327
14	3.6 cm	101,161,492	4,865,950	830
15	3.5 cm	99,753,195	4,515,076	613
16	3.1 cm	96,330,374	5,101,702	873
17	2.8 cm	84,276,897	4,614,972	1197
18	2.7 cm	80,542,538	4,035,966	270
19	2.0 cm	61,707,364	3,858,269	1472
20	2.1 cm	66,210,255	3,439,621	544
21	1.6 cm	45,090,682	2,049,697	234
22	1.7 cm	51,324,926	2,135,311	488
X	5.3 cm	154,259,566	5,753,881	842
Y	2.0 cm	62,460,029	211,643	71
mtDNA	5.4 *µm*	16,569	929	13

A drug database can be used annotate each variant.

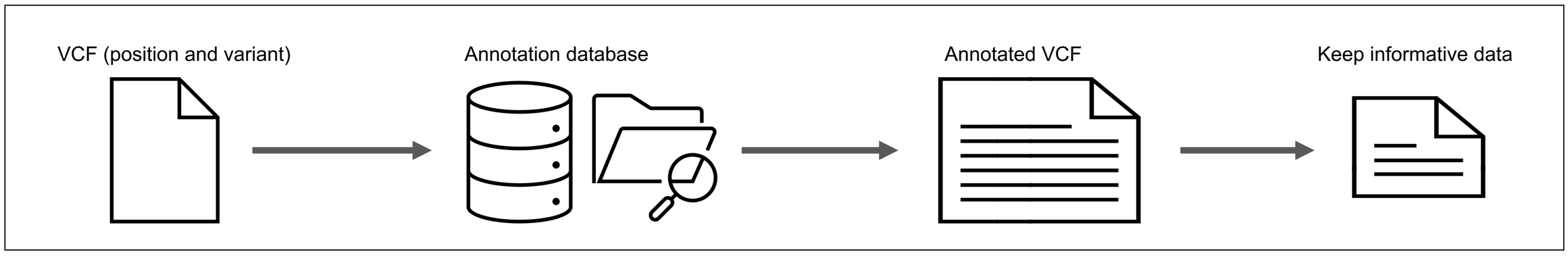
Drug – gene database

Gene	Drug_id	Name	Drug Group	Actions
RAG1	Db00107	Oxytocin	Approved	Binder

After variant effect database + drug – gene database

Position	Ref	Alt	Gene	Consequence	Domain	Clinical_Significance	Drug_id	Name	Drug Group	Action	Interpretation
Chr 1: 1	A	-	RAG1	None	Unknown	Unknown	Db00107	Oxytocin	Approved	Binder	OK
Chr 1: 2	T	-	RAG1	None	Unknown	Unknown	Db00107	Oxytocin	Approved	Binder	OK
Chr 1: 3	G	-	RAG1	None	Catalytic	Benign	Db00107	Oxytocin	Approved	Binder	OK
Chr 1: 4	C	-	RAG1	None	Binding	Pathogenic	Db00107	Oxytocin	Approved	Binder	OK
…											
Chr 1:11	C	A	RAG1	missense	Unknown	Benign	Db00107	Oxytocin	Approved	Binder	Warning
…											
Chr 1: 21	T	G	RAG1	stop	Binding	Pathogenic	Db00107	Oxytocin	Approved	Binder	Warning !!

Summary



Note: Genetic coordinates, Gene, Drug, ID, interpretation are fake – toy data for illustration only.