

Name: _____

Mapping genes to T zone lymphoma in Golden Retrievers using SNPs

For this exercise, you will examine the data to measure which SNPs have the strongest correlation with T zone lymphoma.

1. For each SNP in the dataset, count the number of times an allele appears. Record the number in the appropriate box in the tables below.
 - Note that if the SNP is homozygous at a particular locus (e.g., TT), you will count the T allele twice. If it is heterozygous (e.g., TC) you will count the T allele once and the C allele once.
2. Calculate the observed and expected frequencies for each SNP
3. Calculate the χ^2 value and p-value

Chr5: 193,864

Observed

Allele	Cancer	No Cancer	Total
A			
C			

Expected

Allele	Cancer	No Cancer
A		
C		

X ² Value	p-value

Chr5: 32,613,940*Observed*

Allele	Cancer	No Cancer	Total
A			
G			

Expected

Allele	Cancer	No Cancer
A		
G		

X ² Value	p-value

Chr5: 32,903,109*Observed*

Allele	Cancer	No Cancer	Total
A			
G			

Expected

Allele	Cancer	No Cancer
A		
G		

X ² Value	p-value

Chr5: 58,799,205*Observed*

Allele	Cancer	No Cancer	Total
A			
G			

Expected

Allele	Cancer	No Cancer
A		
G		

X ² Value	p-value

Chr12: 822,209

Observed

Allele	Cancer	No Cancer	Total
A			
G			

Expected

Allele	Cancer	No Cancer
A		
G		

X ² Value	p-value

Chr12: 1,453,973

Observed

Allele	Cancer	No Cancer	Total
A			
G			

Expected

Allele	Cancer	No Cancer
A		
G		

X ² Value	p-value

Chr12: 16,818,273

Observed

Allele	Cancer	No Cancer	Total
A			
G			

Expected

Allele	Cancer	No Cancer
A		
G		

X ² Value	p-value

Chr28: 34,241,635

Observed

Allele	Cancer	No Cancer	Total
A			
G			

Expected

Allele	Cancer	No Cancer
A		
G		

X ² Value	p-value

Questions

1. Which SNP shows the strongest correlation with T zone lymphoma?
2. Which SNP is the least strongly correlated with T zone lymphoma?
3. Which SNPs are statistically significant at the 5% significance level?
4. If you were to look for a gene involved in T zone lymphoma, on which region of the genome would you focus your analysis?

ID	Diagnosis	Chr5: 193864		Chr5: 32613940		Chr5: 32903109		Chr5: 58799205		Chr12: 822209		Chr12: 1453973		Chr12: 16818273		Chr28: 34241635	
		Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2
34911	Case	A	A	A	A	A	G	A	G	A	G	G	G	G	G	G	G
34912	Case	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G
35517	Case	A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	G
35888	Case	A	C	G	G	G	G	A	G	G	G	G	G	G	G	G	G
35889	Case	A	C	G	G	A	G	A	G	A	G	A	G	G	G	G	G
36129	Case	A	A	A	G	A	A	G	G	G	G	G	G	G	G	A	G
37058	Case	A	A	A	A	A	A	A	G	G	G	G	G	G	G	A	G
37209	Case	A	A	G	G	A	A	G	G	G	G	G	G	G	G	A	G
37825	Case	A	A	A	G	G	G	A	G	A	G	G	G	G	G	G	G
37942	Case	A	A	G	G	A	G	A	G	A	G	G	G	G	G	G	G
38116	Case	A	C	A	G	A	A	G	G	A	G	A	G	G	G	A	A
38720	Case	A	A	A	G	A	G	A	A	G	G	G	G	G	G	G	G
39465	Case	A	A	G	G	G	G	A	G	G	G	G	G	G	G	G	G
39785	Case	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G
39961	Case	A	A	A	G	A	A	A	G	A	G	A	G	G	G	A	G
40298	Case	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G
40443	Case	A	A	A	G	G	G	A	G	G	G	G	G	G	G	G	G
40611	Case	A	C	A	G	A	A	G	G	G	G	G	G	G	G	G	G
40695	Case	A	A	A	A	A	A	G	G	A	A	A	A	G	G	A	G
40697	Case	A	C	A	A	G	G	A	G	A	G	G	G	G	G	G	G
41030	Case	A	A	G	G	A	A	A	G	A	G	A	G	G	G	G	G
41032	Case	A	A	G	G	A	A	G	G	G	G	G	G	G	G	G	G
41296	Case	A	A	G	G	G	G	A	G	G	G	G	G	G	G	G	G
41405	Case	A	A	A	G	A	G	A	G	G	G	G	G	G	G	A	G
41636	Case	A	A	G	G	A	G	A	G	G	G	G	G	G	G	G	G
41668	Case	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G
42188	Case	C	C	A	G	A	G	A	A	A	G	G	G	G	G	A	G
42190	Case	A	A	G	G	A	G	A	G	G	G	G	G	G	G	G	G
42299	Case	A	C	A	G	A	A	A	G	G	G	G	G	G	G	G	G
42345	Case	A	A	A	G	A	A	A	G	G	G	G	G	G	G	A	G
42432	Case	A	A	A	A	A	G	A	G	G	G	G	G	G	G	A	G
42447	Case	A	A	A	G	A	A	A	G	A	G	G	G	G	G	G	G
42688	Case	A	C	A	G	A	A	A	G	A	G	A	G	G	G	A	G
42917	Case	A	C	A	A	A	A	A	G	A	G	A	G	G	G	G	G

ID	Diagnosis	Chr5: 193864		Chr5: 32613940		Chr5: 32903109		Chr5: 58799205		Chr12: 822209		Chr12: 1453973		Chr12: 16818273		Chr28: 34241635	
		Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2
43271	Case	A	C	A	G	G	G	G	G	G	G	A	G	A	G	G	G
43337	Case	A	A	A	A	A	G	A	G	G	G	G	G	G	G	G	G
43538	Case	A	A	A	G	A	G	A	A	G	G	G	G	G	G	A	A
43861	Case	A	A	A	G	A	A	G	G	G	G	G	G	G	G	A	G
44041	Case	A	A	A	G	A	G	G	G	G	G	G	G	G	G	A	G
44322	Case	A	A	A	G	A	G	A	G	G	G	G	G	G	G	G	G
44417	Case	A	C	A	G	A	G	G	G	G	G	G	G	G	G	G	G
44666	Case	A	A	A	A	A	A	G	G	A	G	G	G	G	G	A	A
44807	Case	A	A	A	G	A	G	G	G	G	G	G	G	G	G	G	G
44917	Case	A	C	G	G	A	A	G	G	G	G	G	G	G	G	G	G
44919	Case	A	A	A	A	A	G	A	A	A	G	A	G	G	G	A	G
45142	Case	A	A	G	G	G	G	A	G	A	G	G	G	G	G	G	G
45355	Case	A	A	G	G	A	G	A	G	A	G	A	G	G	G	G	G
45425	Case	A	A	A	G	A	A	A	G	A	G	A	G	G	G	G	G
45570	Case	A	A	A	G	A	A	A	G	G	G	G	G	G	G	G	G
45606	Case	A	A	A	G	G	G	A	A	A	G	A	G	G	G	G	G
45744	Case	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G
45868	Case	A	A	G	G	G	G	A	A	G	G	G	G	G	G	G	G
46212	Case	A	C	G	G	A	A	A	G	A	G	G	G	G	G	G	G
46214	Case	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G
46361	Case	C	C	G	G	G	G	A	A	G	G	G	G	G	G	G	G
46386	Case	A	A	A	G	A	A	A	G	G	G	G	G	G	G	G	G
46392	Case	A	A	A	G	G	G	A	G	A	G	A	G	G	G	G	G
46396	Case	A	A	A	G	A	A	G	G	A	A	A	A	G	G	A	G
46774	Case	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G
46881	Case	A	A	A	G	A	A	A	G	A	G	A	G	G	G	A	G
47295	Case	A	C	A	A	A	A	A	A	G	G	G	G	G	G	G	G
47437	Case	A	C	A	G	G	G	G	G	G	G	G	G	G	G	G	G
47634	Case	A	A	A	G	A	G	A	G	A	G	G	G	G	G	G	G
48005	Case	A	A	G	G	A	G	G	G	A	G	G	G	G	G	A	G
48224	Case	A	A	A	A	A	G	A	A	G	G	G	G	G	G	G	G
48310	Case	A	A	A	A	A	A	A	G	G	G	G	G	G	G	A	G
48343	Case	A	C	A	A	A	A	A	G	G	G	G	G	G	G	G	G
48412	Case	A	A	G	G	A	G	A	A	A	G	G	G	G	G	G	G

ID	Diagnosis	Chr5: 193864		Chr5: 32613940		Chr5: 32903109		Chr5: 58799205		Chr12: 822209		Chr12: 1453973		Chr12: 16818273		Chr28: 34241635	
		Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2
47560	Control	A	A	G	G	G	G	A	G	A	G	A	G	G	G	A	G
47569	Control	A	A	G	G	G	G	A	A	G	G	G	G	G	G	G	G
47708	Control	A	A	G	G	A	G	A	G	A	G	G	G	G	G	G	G
47832	Control	A	C	A	G	G	G	A	A	G	G	G	G	G	G	G	G
47857	Control	A	A	A	G	A	A	A	G	G	G	G	G	G	G	A	G
47957	Control	A	A	G	G	G	G	A	G	A	G	G	G	G	G	G	G
48225	Control	A	A	G	G	G	G	G	G	A	G	A	G	G	G	G	G
48227	Control	A	C	A	G	G	G	A	G	G	G	G	G	G	G	G	G
48339	Control	A	A	A	G	G	G	A	G	G	G	G	G	G	G	G	G
48360	Control	A	A	A	G	G	G	A	A	G	G	G	G	G	G	G	G
48441	Control	A	A	A	G	A	G	A	G	G	G	G	G	A	G	G	G
48497	Control	A	A	G	G	G	G	A	G	A	G	G	G	G	G	G	G
48800	Control	A	A	G	G	G	G	A	G	G	G	G	G	G	G	G	G
48894	Control	A	A	G	G	G	G	A	G	A	G	G	G	G	G	G	G
49157	Control	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G
49220	Control	A	A	A	G	A	A	G	G	G	G	G	G	G	G	G	G
49353	Control	A	A	A	G	A	A	A	G	A	G	G	G	G	G	G	G
49548	Control	A	A	A	A	A	G	A	A	A	G	G	G	G	G	G	G
49603	Control	A	C	A	G	G	G	A	G	G	G	G	G	G	G	G	G
49683	Control	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G
49706	Control	A	C	A	G	A	G	A	G	G	G	G	G	G	G	G	G
49708	Control	A	C	G	G	A	A	A	A	G	G	G	G	G	G	G	G
49710	Control	A	A	G	G	A	G	G	G	G	G	G	G	G	G	G	G
49769	Control	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G
49861	Control	A	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G
50012	Control	A	A	A	G	G	G	A	A	G	G	G	G	G	G	G	G
50055	Control	A	A	G	G	A	G	A	A	G	G	G	G	G	G	G	G
50100	Control	A	A	A	G	A	A	A	G	A	G	A	G	G	G	G	G
50124	Control	A	A	G	G	G	G	A	G	G	G	G	G	G	G	G	G
50404	Control	A	C	A	G	A	A	A	G	A	G	G	G	G	G	G	G
50574	Control	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G
50728	Control	A	A	G	G	G	G	A	A	G	G	G	G	G	G	G	G
50765	Control	A	A	A	G	A	G	A	A	A	G	A	G	G	G	G	G
		143	104	84	24	82	26	105	32	40	3	22	1	3	0	10	0

ID	Diagnosis	Chr5: 193864		Chr5: 32613940		Chr5: 32903109		Chr5: 58799205		Chr12: 822209		Chr12: 1453973		Chr12: 16818273		Chr28: 34241635	
		Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2
		2	41	61	121	63	119	40	113	105	142	123	144	141	144	135	145