

Colorectal cancer



1057 cases
960 controls

550K SNPs

APRIL 10

Colorectal cancer data from rs6983267

Table 1 Risk of colorectal neoplasia associated with the 8q24 SNP rs6983267

Panel	Group	Total	Genotype			Frequency	
			GG	GT	TT	G	T
A	All affected individuals	1,027 ^a	352	486	189	0.579	0.421
	Cancers only	620	202	302	116	0.571	0.431
	Adenomas only	407 ^a	150	184	73	0.595	0.405
	Controls	960	235	471	254	0.490	0.510
B	Colorectal cancer	1,027	1,324	2,216	911	0.579	0.421

1027 Colorectal cancer

960 controls

Cancer: 0.57G 0.43T

controls: 0.49G 0.51T

Table 1 Risk of colorectal neoplasia associated with the 8q24 SNP rs6983267

Panel	Group	Total	Genotype			Frequency	
			GG	GT	TT	G	T
A	All affected individuals	1,027 ^a	352	486	189	0.579	0.421
	Cancers only	620	202	302	116	0.569	0.431
	Adenomas only	407 ^a	150	184	73	0.595	0.405
	Controls	960	235	471	254	0.490	0.510
B	Colorectal cancers	4,361	1,324	2,216	821	0.560	0.440

Are these different?

Cancer: 0.57G 0.43T

controls: 0.49G 0.51T

Chi squared

Chi squared

<http://www.graphpad.com/quickcalcs/chisquared1.cfm>



1. [Select category](#)

2. Choose calculator

3. Enter data

4. View res

Compare observed and expected frequencies

This calculator compares observed and expected frequencies with the chi-square test. [Read an example with explanation.](#)

Note that the chi-square test is more commonly used in a very different situation -- to analyze a contingency table. This is appropriate when you wish to compare two or more groups, and the outcome variable is categorical. For example, compare number of patients with postoperative infections after two kinds of operations. If you need to analyze a contingency table, do not use this table. If you have two groups (rows) and two outcomes, use [this calculator](#). If your table is larger, try the free demos of [GraphPad InStat](#) (basic statistics only) and [GraphPad Prism](#) (statistics, nonlinear regression and scientific graphics).

Enter the names of the categories into the first column (optional). Enter the actual number of objects or individuals or events observed in the second column. Then enter the expected number, fraction or percent expected in the third column.

1. Choose data entry format

- Enter up to 20 categories (rows).
- Enter or paste up to 2000 categories (rows).

Caution: Changing format will erase your data.

2. How will you enter the expected values?

- Actual number expected
- Percent expected
- Fraction expected

3. Enter data

	Category	Observed #	Expected
1:	<input type="text"/>	<input type="text"/>	<input type="text"/>
2:	<input type="text"/>	<input type="text"/>	<input type="text"/>
	<input type="text"/>	<input type="text"/>	<input type="text"/>

4. View the results

Chi squared

<http://www.graphpad.com/quickcalcs/chisquared1.cfm>

1. Choose data entry format

- Enter up to 20 categories (rows).
- Enter or paste up to 2000 categories (rows).

Caution: Changing format will erase your data.

2. How will you enter the expected values?

- Actual number expected
- Percent expected
- Fraction expected

3. Enter data

	Category	Observed #	Expected
1:	G alleles	1190	.49
2:	T alleles	864	.51
3:			

4. View the results

Calculate now

Clear the form

Chi squared = 31

P values = 10^{-7}

Stuart's genotype

search your account

Go

stuart kim

Acco

atics just got personal.

browse raw data

Showing raw data for SNP **rs6983267**, which is on chromosome **8**.

8
146M Bases
989 Genes
33k SNPs

Jump to a gene:

Go

a SNP: rs6983267

Go

or a chromosome:

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17


18

19

20

21

« Return to your whole genome.

Gene	Position	SNP	Versions	stuart kim's Genotype
 <i>intergenic</i>	128482487	rs6983267	G or T	GG



Homozygous bad allele 😞

Other models

Table 1 Risk of colorectal neoplasia associated with the 8q24 SNP

Panel	Group	Total	Genotype		
			GG	GT	TT
A	All affected individuals	1,027 ^a	352	486	189
	Cancers only	620	202	302	116
	Adenomas only	407 ^a	150	184	73
	Controls	960	235	471	254
B	Colorectal cancers	436	133	216	87

Dominant: Assume G is dominant.
GG or GT vs TT

	GG or GT	TT
Cases	838	189
Controls	706	254

Other models

Table 1 Risk of colorectal neoplasia associated with the 8q24 SNP

Panel	Group	Total	Genotype		
			GG	GT	TT
A	All affected individuals	1,027 ^a	352	486	189
	Cancers only	620	202	302	116
	Adenomas only	407 ^a	150	184	73
	Controls	960	235	471	254
B	Colorectal cancers	4,361	1,334	2,216	811

Recessive: Assume G is recessive.
GG vs GT or TT

	GG	GT or TT
Cases	352	675
Controls	235	725

Other models

Table 1 Risk of colorectal neoplasia associated with the 8q24 SNP

Panel	Group	Total	Genotype		
			GG	GT	TT
A	All affected individuals	1,027 ^a	352	486	189
	Cancers only	620	202	302	116
	Adenomas only	407 ^a	150	184	73
	Controls	960	235	471	254
B	Colorectal cancers	4,361	1,334	2,216	811

additive: GG > GT > TT

Do linear regression 3 genotype x 2 groups

How different is this SNP in the cases versus the controls?

Allelic odds ratio: ratio of the allele ratios in the cases divided by the allele ratios in the controls

$$\text{CASES} \quad .57 \text{ G} / .43 \text{ T} = 1.32$$

$$\text{---} = 1.37$$

$$\text{CONTROLS} \quad .49 \text{ G} / .51 \text{ T} = 0.96$$

How different is this SNP in the cases versus everyone?

Allelic odds ratio*: ratio of the allele ratios in the cases divided by the allele ratio in the entire population

(need allele ratio from entire population to do this)

Likelihood ratio: What is the likelihood of seeing a genotype given the disease compared to the likelihood of seeing the genotype given no disease?

(need data from entire population to do this. We can do this in the class GWAS. For cancer vs controls, the two groups were separate and so we do not know the genotype frequencies of the population as a whole.)

Increased Risk: What is the likelihood of seeing a trait given a genotype compared to overall likelihood of seeing the trait in the population?

(need data from entire population to do this. We can do this in the class GWAS. For cancer vs controls, the two groups were separate and so we do not know the genotype frequencies of the population as a whole.)

Multiple hypothesis testing

“Of the 547,647 polymorphic tag SNPs, 27,673 showed an association with disease at $P < .05$.”

- $P = .05$ means that there is a 5% chance for this to occur randomly.
- If you try 100 times, you will get about 5 hits.
- If you try 547,647 times, you should expect $547,647 \times .05 = 27,382$ hits.
- So 27,673 (observed) is about the same as one would randomly expect.

Multiple hypothesis testing

“Of the 547,647 polymorphic tag SNPs, 27,673 showed an association with disease at $P < .05$.”

- Here, have 547,647 SNPs = # hypotheses
- False discover rate = $q = p \times \# \text{ hypotheses}$. This is called the Bonferroni correction.
- Want $q = .05$. This means a positive SNP has a .05 likelihood of rising by chance.
- At $q = .05$, $p = .05 / 547,647 = .91 \times 10^{-7}$
- This is the p value cutoff used in the paper.

Multiple hypothesis testing

“Of the 547,647 polymorphic tag SNPs, 27,673 showed an association with disease at $P < .05$.”

- The Bonferroni correction is too conservative. It assumes that all of the tests are independent.
- But the SNPs are linked in haplotype blocks, so there really are less independent hypotheses than SNPs.
- Another way to correct is to permute the data many times, and see how many times a SNP comes up in the permuted data at a particular threshold.

SNPedia

Summarize the trait

Summarize the study

- How large was the cohort?

- How strong was the p-value?

- What was the OR, likelihood ratio or increased risk?

Which population?

What is known about the SNP?

- Associated genes?

- Protein coding?

- Allele frequency?

Does knowledge of the SNP affect diagnosis or treatment?