Bio84
Homework Assignment 3
2/17/2012

Assignment: Describe a Genome-Wide Association Study for an inherited disease or condition

Kidney Stones
Having suffered kidney stone attacks on the average of once per decade for the last 40 years, three of which sent me in excruciating pain to my local emergency room, I was naturally interested to see if there has been a GWAS study of this condition. My brother has endured at least one stone, and I am told that one of my three maternal uncles had such problems with stones that he had a kidney removed to end his suffering. But neither my mother nor my father had stones. So is this an inherited condition? Perhaps. Let's see what the literature offers.

The GWAS
Only one study was referenced in the NHGRI catalog of GWAS:
http://www.genome.gov/GWAStudies/index.cfm?pageid=26525384#searchForm

Here is the summary of the study as published in the catalog:

Date Added to Catalog (since 11/25/08)	First Author/Date/ Journal/Study	Disease/ Trait	Initial Sample Size	Replication Sample Size	Region	Reported Gene(s)	Mapped Gene(s)	Strongest SNP-Risk Allele	Context	Risk Allele Frequency in Controls	P-value	OR or beta-coefficient and [95% CI]	Platform	CNV
7/1/2009	Thorleifsson G 28-Jun-09 *Nat Genet* Sequence variants in the CLDN14 gene associate with kidney stones and bone mineral density.	Kidney stones	1,507 Icelandic cases, 34,033 Icelandic controls	1,520 Icelandic cases, 4,726 Icelandic controls, 746 Dutch cases, 3,751 Dutch controls	21q22.13	*CLDN14*	CLDN14	rs219780-C	cds-synon	0.79	4 x 10-12	1.25 [1.17-1.33]	Illumina [303,120]	N

As we can see, this was a reasonably sized study, performed on two northern European populations, and with results implicating a particular gene. This looked promising.

After reading the abstract in PubMed:
http://www.ncbi.nlm.nih.gov/pubmed/19561606?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum
I was motivated to read the entire study, which I purchased at:
http://www.nature.com/doifinder
I have attached a copy of the downloaded study to the e-mail submitting this homework assignment.

Summary of the Study

The study was conducted by an international team headed by Kari Stefansson of Reykjavik University. They assembled samples from 3,773 kidney stone cases and 42,510 control subjects in Iceland and The Netherlands. Several SNPs on chromosome 21q22 were discovered to be strongly associated with nephrolithiasis (that's the fancy name for kidney stones). As a result, the experiment was repeated on both Icelandic and Dutch patients, and the results confirmed the original finding.

Two SNPs, rs 219778 and rs219781 were initially strongly associated with kidney stones:

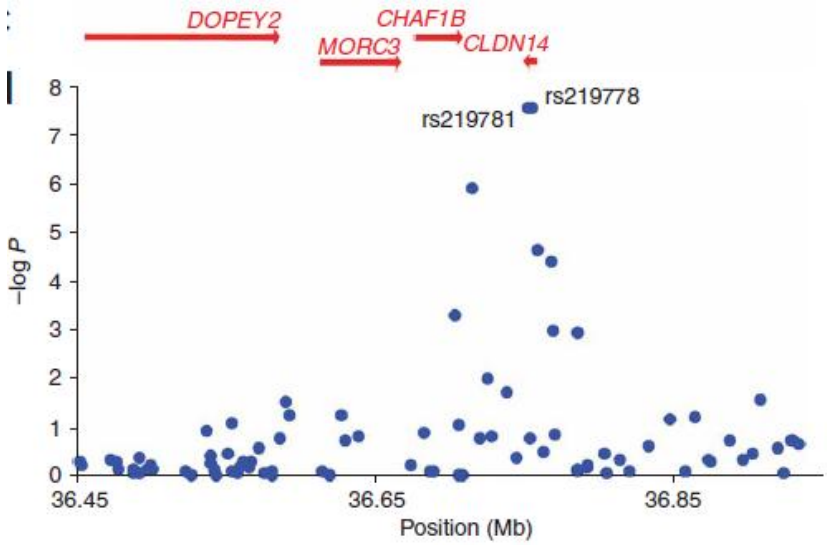


Figure 1 from Thorleifsson et al. The two strongly associated SNPs lie on either side of the last exon of the CLDN14 (claudin 14) gene.

These two SNPs, though non-exonic themselves, lie on either side of the last exon of a gene known to code for a *claudin* protein that controls permeability of epithelial cells. The trail was getting warm! These two SNPs were associated with an odds ratio of 1.3 for kidney stones [ref. Supplementary Table 1], but the fact that both were associated turned out to be due entirely to their close proximity. The researchers used rs219781 as a reference and, after controlling for linkage to other SNPs in the locus, found only rs219781 to have a significant odds ratio [OR=1.3, $P=3.2 \times 10^{-8}$] for kidney stones.

In an effort to find the true culprit (rs219781 had a perfect alibi: it lived in an intron), the team searched nearby. They found two synonymous SNPs rs217979 and rs217980, both of which lie in "the last and only translated exon of CLDN14", determined their genotypic variants, and associated these with kidney stones. The results were dramatic, as shown in their Supplementary Table 4:

Supplementary Table 4: Genotype specific odds ratios for the risk alleles of rs219780 and rs219779

Shown is the risk for heterozygous carriers (CT) and homozygous carriers (CC) of rs219780 and rs219779 compared to the risk for non-carriers (TT), together with 95% confidence intervals (CI), both for the combined Icelandic discovery and replication sample set and the sample set from the Netherlands, and for the sample sets combined using a Mantel-Haenzel model. For the Icelandic sample sets the *P* values are adjusted for relatedness using simulations.

Cohort (N_C/N_A)		**Genotype specific Odds Ratio**	
Variant AT	**TT**	**CT (95% CI)**	**CC (95% CI)**
Iceland Combined (38759/3027)			
rs219780	1	1.24 (0.99-1.54)	1.58 (1.29-1.95)
rs219779	1	1.24 (1.04-1.50)	1.56 (1.31-1.86)
The Netherlands (3751/746)			
rs219780	1	1.81 (1.14-2.87)	1.96 (1.24-3.08)
rs219779	1	1.36 (0.96-1.93)	1.52 (1.07-2.16)
Combined (42510/3773)			
rs219780	1	1.33 (1.09-1.62)	1.64 (1.36-1.98)
rs219779	1	1.27 (1.08-1.49)	1.55 (1.33-1.82)

This table clearly shows that when the normal allele (TT) becomes either heterozygous (CT) or homozygous (CC) for T>>C, the odds ratio increases dramatically. Individuals with the rs219780 CC allele can expect between 36% and 98% increase in the odds of getting a kidney stone, and for rs219779, between 33% and 82% increase. [Red box highlight added].

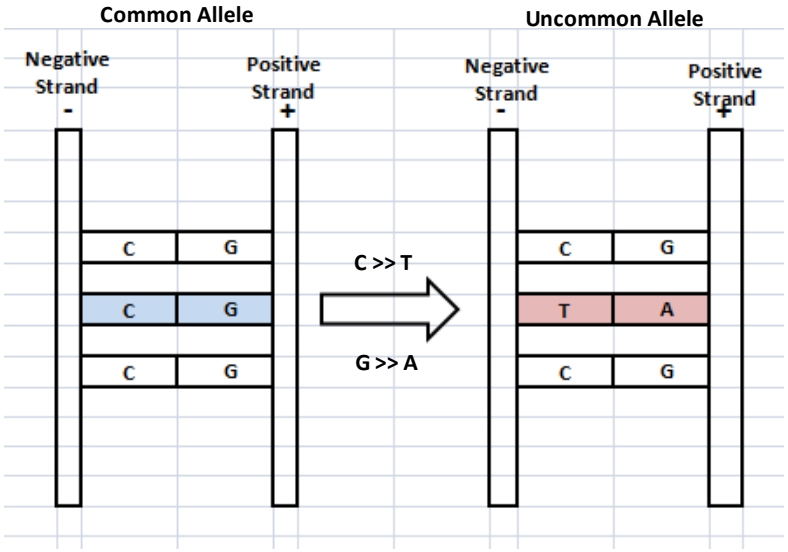
Am I, with my suspected family history of kidney stones, the proud owner of one of those CT or CC alleles? I looked forward to finding out from 23andMe. But researching their website reveals they seem to have ignored this paper in favor of another one from Stefansson's group, a paper dealing primarily with chronic kidney disease, and only secondarily with kidney stones:
https://www.23andme.com/health/kidney-stones/
Gudbjartsson DF et al. (2010) . "Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases." *PLoS Genet* 6(7):e1001039.

When I read this paper online at:
http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1001039

I found that the SNP associated with chronic kidney disease actually had a reduced odds ratio for kidney stones [OR= 0.88 for rs4293393-T]. 23andMe report an odds ratio of 1.14. I have e-mailed them in an attempt to clarify this issue and to see whether they can test for rs219779-80.

Uh-Oh!

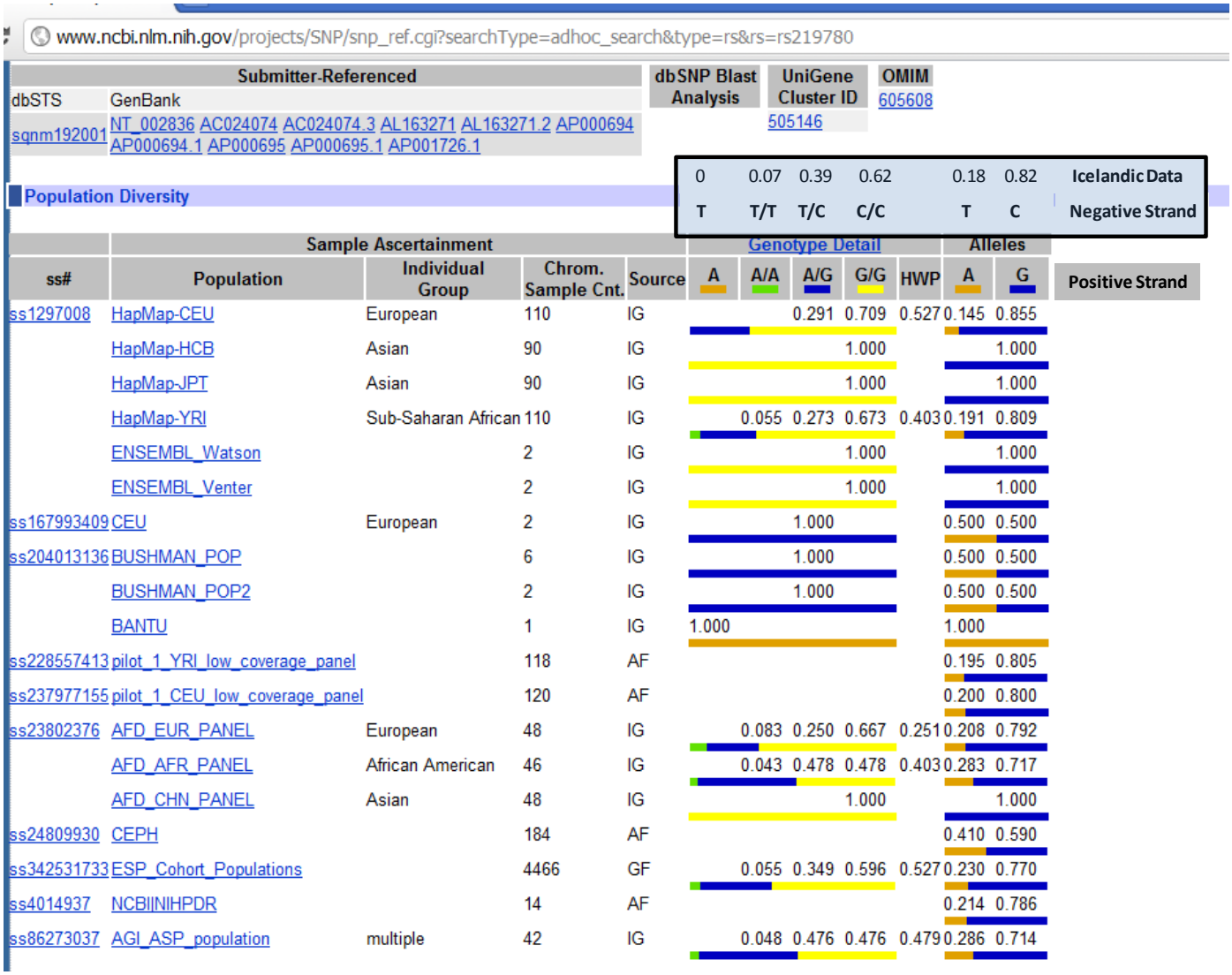
I received a reply from 23andMe which set me straight on some misconceptions about the meaning of the reported results in the Thorleifsson paper. To understand the issue, we first need to understand that there are two different ways of analyzing and reporting SNP data in the literature. It turns out there are two different, but consistent, ways of specifying results, almost as if they were reported in two different languages.



Genotyping results are specified as having been read from either the *Positive Strand* or the *Negative Strand* of the DNA. When a SNP mutates from C to T on the negative strand, it mutates from G to A on the positive strand. Authors are not always clear about which strand they are using. But once this is understood, everything becomes clear.

Second, and most important for the article being reviewed, is the need to understand the relative frequency of occurrence of the SNPs reported: which is the common allele and which is the uncommon one? Careful re-reading of the Thorleifsson paper showed which was which, but a much better way to find this information is through dbSNP:
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?searchType=adhoc_search&type=rs&rs=rs219780

The results are shown on the next page.



www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?searchType=adhoc_search&type=rs&rs=rs219780

Submitter-Referenced		dbSNP Blast Analysis	UniGene Cluster ID	OMIM
dbSTS	GenBank			
sqnm192001	NT_002836 AC024074 AC024074.3 AL163271 AL163271.2 AP000694 AP000694.1 AP000695 AP000695.1 AP001726.1		505146	605608

0	0.07	0.39	0.62	0.18	0.82	Icelandic Data
T	T/T	T/C	C/C	T	C	Negative Strand

Population Diversity

	Sample Ascertainment				Genotype Detail					Alleles		
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	A	A/A	A/G	G/G	HWP	A	G	Positive Strand
ss1297008	HapMap-CEU	European	110	IG			0.291	0.709	0.527	0.145	0.855	
	HapMap-HCB	Asian	90	IG				1.000			1.000	
	HapMap-JPT	Asian	90	IG				1.000			1.000	
	HapMap-YRI	Sub-Saharan African	110	IG		0.055	0.273	0.673	0.403	0.191	0.809	
	ENSEMBL_Watson		2	IG				1.000			1.000	
	ENSEMBL_Venter		2	IG				1.000			1.000	
ss167993409	CEU	European	2	IG			1.000			0.500	0.500	
ss204013136	BUSHMAN_POP		6	IG			1.000			0.500	0.500	
	BUSHMAN_POP2		2	IG			1.000			0.500	0.500	
	BANTU		1	IG	1.000					1.000		
ss228557413	pilot_1_YRI_low_coverage_panel		118	AF						0.195	0.805	
ss237977155	pilot_1_CEU_low_coverage_panel		120	AF						0.200	0.800	
ss23802376	AFD_EUR_PANEL	European	48	IG		0.083	0.250	0.667	0.251	0.208	0.792	
	AFD_AFR_PANEL	African American	46	IG		0.043	0.478	0.478	0.403	0.283	0.717	
	AFD_CHN_PANEL	Asian	48	IG				1.000			1.000	
ss24809930	CEPH		184	AF						0.410	0.590	
ss342531733	ESP_Cohort_Populations		4466	GF		0.055	0.349	0.596	0.527	0.230	0.770	
ss4014937	NCBI|NIHPDR		14	AF						0.214	0.786	
ss86273037	AGI_ASP_population	multiple	42	IG		0.048	0.476	0.476	0.479	0.286	0.714	

I have superimposed the results as reported in Thorleifsson above the table of relative frequency of occurrence of the alleles [C/T or G/A] and the Genotypes [C/C, T/C, T/T; G/G, A/G, A/A] from each of the studies reported in the literature.[1] Notice that the (negative strand) allele C, the one which is correlated with a high odds ratio for kidney stones relative to the allele T, is actually by far the most common in most populations. In fact the Asian samples are 100% homozygous for kidney stone risk. The Europeans are 50% - 70% homozygous for stones and risky allele [C] occurs in over 80% of Europeans. So while it is true, as reported, that the odds ratio for CC compared to TT is high, signaling a risk of kidney stones, it is NOT true, as I reported above, that TT is the common allele. It is not. As a result, only a lucky few of us will be in possession of the benign allele of rs219780, and therefore less likely than our fellows to suffer the excruciating pain of nephrolithiasis.

[1] ss= “submitted SNP”. Once accepted, it becomes rs=”reference SNP”