An example of following up results in a two-stage GWAS design

David Duffy

In a 100K Affymetrix-based GWAS we carried out in 2006-7, one SNP on chromosome 19 exhibited particularly strong association to our phenotype of interest, nevus count.

The original result causing such excitement (275 usable obs). The P-value ignores the relatedness of the twins.

```
                      Moles
rs3848620       Mean              SD       Count
           ---------------------------------------
    1/2          108.2075        48.9109      53
    2/2           73.4730        41.3354     222

Kruskal-Wallis H=  24.07 df=  1 (P=0.0000)
```

We were slightly cautious about this accepting this result, because none of the neighbouring SNPs gave any evidence of association to nevus count. However, this SNP was not in strong LD with any neighbours. In addition, this SNP has been removed from later Affymetrix arrays, presumably because of assay problems, but no information was available.

Therefore, we carried out additional genotyping on the Sequenom platform.

```
Marker         NAll  Allele(s)    Freq   Het    Ntyped
-------------- ---- -----------   ------ ------ ------
rs3848620         2     1     (2) 0.1079 0.1928    329  Original
rs3848620F        2     A     (G) 0.3515 0.4560   3799  Forward
rs3848620R        2     A     (G) 0.3539 0.4574   3745  Reverse
```

We found the minor allele frequency to be quite different from the original genotyping.

Crosstabulating the old and new assays, the Affymetrix assay is adequate for the G/G homozygotes, but the A/A and A/G were pretty bad.

```
                      rs3848620F
rs3848620     A/A          A/G          G/G
          -------------------------------------
    1/2    45 (.662)    23 (.338)      0 (.000)
    2/2     0 (.000)    58 (.249)    175 (.751)
```

Even the two different Sequenom assays disagreed with one another in a few cases - the *rs3848620R* assay also gave more Mendelian errors. One would presume a proportion of true A/G heterozygotes are being called as A/A by the reverse strand assay. This is a classical pattern of assay error.

Even so, the overall agreement was 99.7%

```
                     rs3848620R
rs3848620F       A/A          A/G         G/G
           -------------------------------------
    A/A      489 (1.00)     0 (.000)    0 (.000)
    A/G       10 (.006)  1653 (.994)    0 (.000)
    G/G        0 (.000)     0 (.000) 1592 (1.00)
```

And, the new assay in the original subsample (250 usable obs) still gave nice evidence of association:

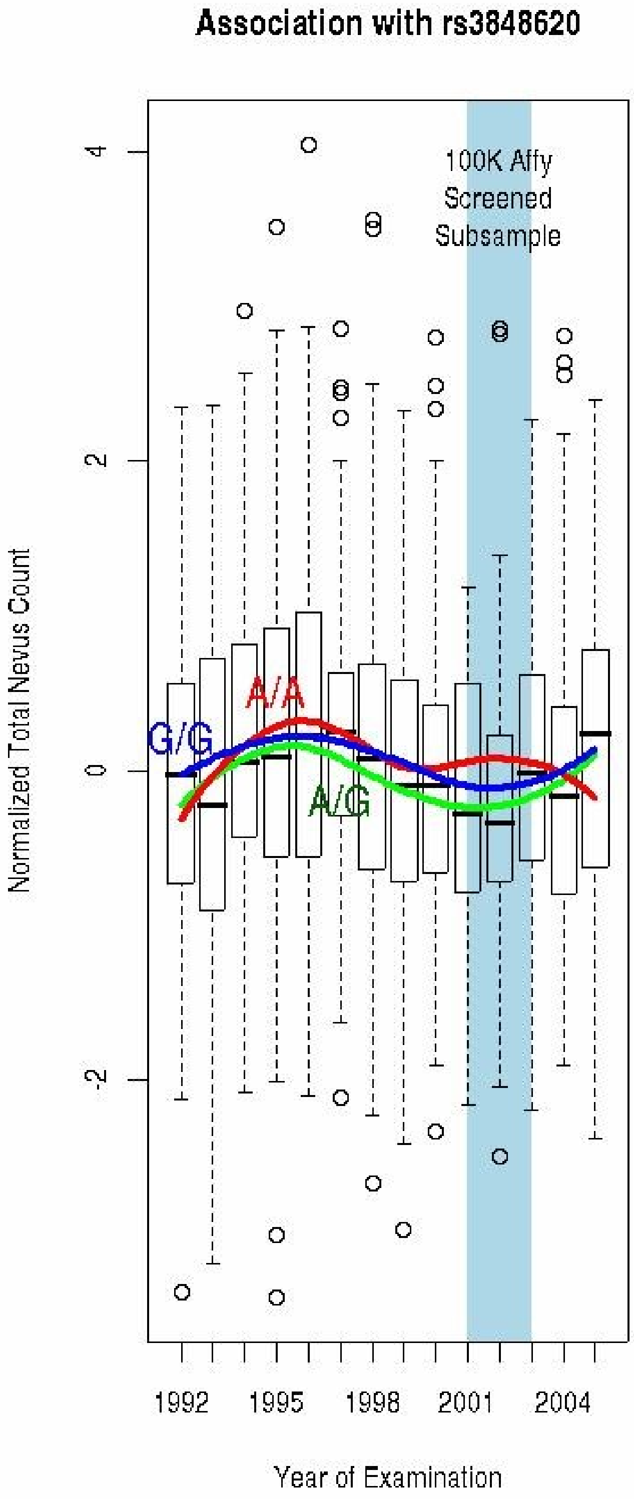
```
                     Moles
rs3848620F      Mean                  SD       Count
           --------------------------------------
    A/A          121.9697        44.4512      33
    A/G           71.8857        46.5345      70
    G/G           75.8707        39.5901     147

Kruskal-Wallis H=  32.08 df=  2 (P=0.0000)
```

But, using all the available observations.

```
                     V1.NF
rs3848620F      Mean              SD       Count
           ---------------------------------------
    A/A           91.7244        53.7296     225
    A/G           83.3560        46.2774     795
    G/G           87.3276        52.5656     760

Kruskal-Wallis H=   3.66 df=  2 (P=0.1605)
```



The figure shows the relationship between genotype, residualized nevus count and year of examination. The Affy 100 K sample were all examined in 2001-2003.