SNPs and Cancer Prognostics

Terrence P McGarty

White Paper No 99
July, 2013

There has been a great deal of work determining the prognostic value of SNPs and even SNP-SNP pairs. We examine this here and do so in a somewhat critical manner. One of the problems is that the analyses are oftentimes not reproducible and in all cases seem to lack any causative models. Copyright 2013 Terrence P. McGarty, all rights reserved.

Notice

This document represents the personal opinion of the author and is not meant to be in any way the offering of medical advice or otherwise. It represents solely an analysis by the author of certain data which is generally available. The author furthermore makes no representations that the data available in the referenced papers is free from error. The Author also does not represent in any manner or fashion that the documents and information contained herein can be used other than for expressing the opinions of the Author. Any use made and actions resulting directly or otherwise from any of the documents, information, analyses, or data or otherwise is the sole responsibility of the user and The Author expressly takes no liability for any direct or indirect losses, harm, damage or otherwise resulting from the use or reliance upon any of the Author's opinions as herein expressed. There is no representation by The Author, express or otherwise, that the materials contained herein are investment advice, business advice, legal advice, medical advice or in any way should be relied upon by anyone for any purpose. The Author does not provide any financial, investment, medical, legal or similar advice in this document or in its publications on any related Internet sites. This document is subject to change, correction, and withdrawal at any time and without notice. Thus there should be no reliance on the data or conclusions contained herein. This is a work in progress and should be used as such.

Contents

1 Introduction ... 3
1.1 SNP a Definition ... 3
1.2 SNP Operations ... 5
2 SNP and Cancers ... 7
2.1 Prostate ... 10
2.2 Melanoma ... 16
3 SNP-SNP Interactions ... 19
3.1 SNP-SNP Analysis ... 21
3.2 Specific Genes ... 25
3.2.1 MMP ... 25
3.2.2 ROBO1 ... 27
3.2.3 CSF1 ... 28
3.2.4 EGFR ... 29
3.2.5 FBLN5 ... 30
3.2.6 HSPG2 ... 31
3.3 The Clustering Approach ... 32
4 Observations ... 36
4.1 Specifics ... 36
4.2 Models and World View ... 38
5 References ... 40
5.1 Web Sites ... 40
5.2 Documents ... 40
6 Abbreviations: ... 44

1 INTRODUCTION

SNPs or single nucleotide polymorphisms are single nucleotide changes from what would be a "normal" pattern in DNA[1]. There are lots of SNPs but to be a true SNP it must be different but not too different. Namely the nucleotide location must be saying T for almost all people but for a small section of people it may become A. SNPs have been identified across the human genome. Furthermore SNPs have been correlated to a multiplicity of cancers. The question is causation or correlation or coincidence. Unlike various proteins expressed by genes the SNPs are a single nucleotide change, and a change from many of the other human genes but not necessarily a unique change.

In this brief section we examine SNPs as applied to two malignancies; prostate cancer and melanoma. Recent research has examined prostate cancer and a collection of other cancers using a genome wide survey approach. It is our opinion that such a wide net may all too often collect relationships which may or may not be causal. We have argued before that having a model of causation, a clear path of cause and effect, and then validating such with in this case a SNP has merit. Just noting relationships between SNP presence and a disease may be mere coincidence.

We review some work on single SNP analyses and SNP-SNP analyses as well. One of our concerns is that the overall analysis is oftentimes highly subjects and not repeatable. SNPs clearly play a role yet the causative nature is not clearly incorporated in the analysis. Thus our current overall acceptance of SNPs for prognostic inference should be highly suspect. We demonstrate that herein.

1.1 SNP A DEFINITION

As is stated in the NIH description:

Single nucleotide polymorphisms, frequently called SNPs (pronounced "snips"), are the most common type of genetic variation among people[2]. Each SNP represents a difference in a single

[1] Single Nucleotide Polymorphisms (SNPs) Polymorphisms defined by SNPs may occur within or outside of a protein coding sequence. If the SNP occurs within a gene, the SNP allele can be designated based on its dbSNP_id, followed by a hyphen and the specific nucleotide.

Examples:

Park2rs6200232-G The Park2 rs6200232 SNP allele with the G variant

Park2rs6200232-A The Park2 rs6200232 SNP allele with the A variant

If the SNP occurs outside of an identified gene, the SNP locus can be designated using the dbSNP_id as the locus symbol and the nucleotide allelic variants are then superscripted as alleles. If a gene is later discovered to include this SNP locus, the same guidelines are applicable as those used when mutant locus symbols become alleles of known genes.

Examples:

rs6200616T A SNP locus with the T variant

rs6200616C A SNP locus with the C variant

Note: If a gene Xyz is later discovered to include this SNP locus, rs620061, then the alleles listed above become Xyzrs620061-T and Xyzrs620061-C. http://www.informatics.jax.org/mgihome/nomen/gene.shtml
\

[2] http://ghr.nlm.nih.gov/handbook/genomicresearch/snp

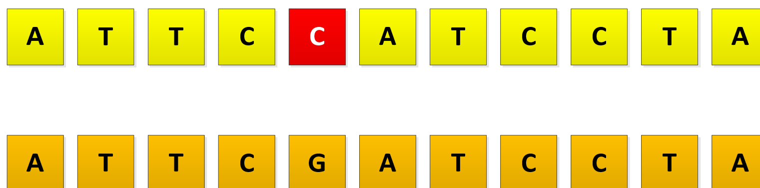
DNA building block, called a nucleotide. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain stretch of DNA.

SNPs occur normally throughout a person's DNA. They occur once in every 300 nucleotides on average, which means there are roughly 10 million SNPs in the human genome. Most commonly, these variations are found in the DNA between genes. They can act as biological markers, helping scientists locate genes that are associated with disease. When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene's function.

Most SNPs have no effect on health or development. Some of these genetic differences, however, have proven to be very important in the study of human health. Researchers have found SNPs that may help predict an individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases. SNPs can also be used to track the inheritance of disease genes within families. Future studies will work to identify SNPs associated with complex diseases such as heart disease, diabetes, and cancer.

In contrast in Speicher et al the authors state that SNPs occur once every 1,000 nucleotides rather than once every 300. They contend that the typical human may have 3 million SNPs. Furthermore a major SNP would have an occurrence of 5% or more. Recall that if we replace a T for a C then we could also have replaced a G or A as well. The dominant nucleotide is the once which is most prevalent.

We demonstrate graphically below an example of a SNP.



A SNP is a single nucleotide change which occurs in 1% or more of the population.

This is a single nucleotide change whose prevalence is low but usually greater than 5%.

From the University of Utah site we also have[3]:

Not all single-nucleotide changes are SNPs, though. To be classified as a SNP, two or more versions of a sequence must each be present in at least one percent of the general population.

[3] http://learn.genetics.utah.edu/content/health/pharma/snips/

SNPs occur throughout the human genome - about one in every 300 nucleotide base pairs. This translates to about 10 million SNPs within the 3-billion-nucleotide human genome.

They continue:

SNPs are divided into two main categories:

1. ***Linked SNPs*** *(also called indicative SNPs) do not reside within genes and do not affect protein function. Nevertheless, they do correspond to a particular drug response or to the risk for getting a certain disease.*

2. ***Causative SNPs*** *affect the way a protein functions, correlating with a disease or influencing a person's response to medication. Causative SNPs come in two forms:*

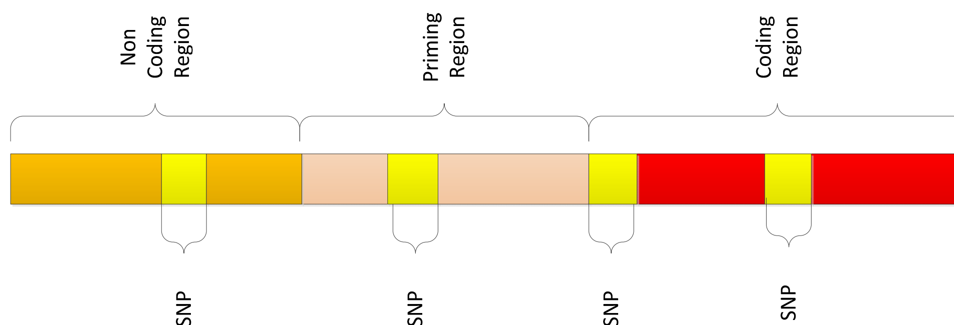
   a) ***Coding SNPs****, located within the coding region of a gene, change the amino acid sequence of the gene's protein product.*

   b) ***Non-coding SNPs****, located within the gene's regulatory sequences, change the level of gene expression and, therefore, how much RNA and protein is produced.*

Thus SNPs may serve many functions depending on where they are and what they have changed to. The question is; are the SNPs germline or somatic or both. Namely are the SNPs always there and they just add the extra push for a malignant cell or does the SNP occur as a part of the overall cell changes as part of its malignancy. Perhaps it is the malignant process which causes the SNP to occur and not the SNP causing the malignancy.

1.2 SNP Operations

SNPs can have multiple actions in the gene. The action will to a degree depend upon where the SNP is. Below we demonstrate three generic regions; non coding region, primer region, and coding region. Generally if the SNP is in the non-coding region one would expect that the impact would be negligible. However this may not necessarily be the case.



Thus depending on where the SNP is located it may or may not have any effect. We have also examined the causes of the SNPs and certain factors such as backscatter X-rays may have

enough power under the right circumstances to induce a SNP in a melanocyte in the basal layer of the skin and thus instigate a melanoma. Likewise one would also see UV light having a similar effect.

As Wakeley et al state:

Single-nucleotide polymorphisms (SNPs) are the markers of choice, both for studies of linkage and for studies of historical demography. This is due to (a) the relative abundance of SNPs in the human genome, compared with other types of polymorphisms, (b) the efficiency with which they can be assayed, and (c) the ease with which they can be analyzed by the tools of population genetics. It is typically assumed that each SNP is the result of a single mutation event and that different SNPs segregate independently of one another. These assumptions are probably correct much of the time.

Then, it is the allele frequencies at SNPs, as well as the distribution of the polymorphisms among subpopulations, that can tell us about demographic history. However, SNPs are discovered—and, later, genotyped—by primer pairs that amplify short fragments of the genome rather than single sites. We refer to these SNP-discovered loci as "SDLs." Some proportion of SDLs will be found to contain multiple SNPs, especially as the sample sizes from human populations increase. This represents an opportunity to garner more information from polymorphism data— namely, the number of SNPs per SDL, denoted by "S," and their joint frequencies in a sample.

2 SNP AND CANCERS

We now will examine SNPs and certain cancers. We have chosen prostate cancer and melanoma because a great deal has been done on both and each represent a prototype of a solid tumor; glandular and epithelial respectively.

SNP analysis has evolved significantly in the past decade. A 2001 paper by Syvanen discussed the technology at the turn of the century. One can see the impact of high throughput techniques beginning to show their value. As she states:

Comparison of genomic DNA sequences in different individuals reveals some positions at which two, or in some cases more than two, bases can occur. These single nucleotide polymorphisms (SNPs) are highly abundant, and are estimated to occur at 1 out of every 1,000 bases in the human genome.

Depending on where a SNP occurs, it might have different consequences at the phenotypic level. SNPs in the coding regions of genes that alter the function or structure of the encoded proteins are a necessary and sufficient cause of most of the known recessively or dominantly inherited monogenic disorders. These SNPs are routinely analysed for diagnostic purposes.

Another important group of SNPs are those that alter the primary structure of a protein involved in drug metabolism. These SNPs are targets for pharmacogenetic analyses. Missense SNPs in the coding regions of genes, such as the two SNPs in the apolipoprotein E gene and the factor V Leiden mutation, can also contribute to common disease. This type of SNP can be analysed to assess the risk of an individual for a particular disease. In addition, it is likely that SNPs in the regulatory regions of genes might influence the risk of common disease.

However, most SNPs are located in non-coding regions of the genome, and have no direct known impact on the phenotype of an individual.

At issue are the SNPs in the non-coding region. Perhaps they may have effects but those effects are yet to be understood. A classic example of SNPs is what we see in the case of blood type, A and B produce protein but O due to a loss of a nucleotide does not thus allowing it to be a donor blood.

In a more recent paper by Grochola et al the authors focus on the p53 gene, a key control element of cell growth, and the impact of SNP changes on that specific pathway. They state:

The p53 tumor suppressor pathway is central both in reducing cancer frequency in vertebrates and in mediating the response of commonly used cancer therapies. This article aims to summarize and discuss a large body of evidence suggesting that the p53 pathway harbors functional inherited single-nucleotide polymorphisms (SNPs) that affect p53 signaling in cells, resulting in differences in cancer risk and clinical outcome in humans.

The insights gained through these studies into how the functional p53 pathway SNPs could help in the tailoring of cancer therapies to the individual are discussed. Moreover, recent work is discussed that suggests that many more functional p53 pathway SNPs are yet to be fully characterized and that a thorough analysis of the functional human genetics of this important tumor suppressor pathway is required.

This study is one of many recent ones which connects SNPs to specific pathway control functions, rather than just trying to align SNPs in some causative role.

In a recent British Journal of Cancer article by Pashayan et al the authors have performed a preliminary analysis of genetic screening of those for higher risk for prostate and breast cancers[4]. We herein look at the prostate cancer issue.

Simply stated the authors have assembled a database of genetic samples and for each have detailed the relative risk and the prevalence. Specifically:

1. They listed SNPs from the dbSNP ("Single Nucleotide Polymorphism database"). A SNP is a DNA sequence variation with a single nucleotide, ATGC, and may be in an exon or intron. Many of these variations occur.

2. The odds ratio, OR, is the odds of an event occurring in one group as compared to another. Thus we can say that if we have two groups, say group 1 which has the SNP alteration, and Group 0 which does not have the alteration, then the odds ratio is given by:

$$[p1/(1-p1)]/[p0/(1-p0)]$$

and if the odds ratio is greater than one then we have a greater chance of occurrence. Now consider two SNPs, and their respective individual and total odds ratio. Let p1 be SNP1 and p2 SNP2 and p0 be the lack of SNP1 and p00 the lack of SNP2. Then we have an odds ratio for both occurring, if independent, as:

$$[p1p2/(1-p1p2)]/[p0p00/(1-p0p00)]$$

This assumes independence and shows that the OR do not readily allow direct and simple calculation from each other separately. We of course can extend this principle to n SNPs. It is obvious

3. Using the SNPs as a measure of increased or decreased risk, one can set a risk threshold and test those above and ignore those below.

The result is given by the authors as:

Compared with screening men based on age alone (aged 55–79: 10-year absolute risk 2%), personalized screening of men age 45–79 at the same risk threshold would result in 16% fewer

[4] http://www.nature.com/bjc/journal/v104/n10/full/bjc2011118a.html#bib28

men being eligible for screening at a cost of 3% fewer screen-detectable cases, but with added benefit of detecting additional cases in younger men at high risk. Similarly, compared with screening women based on age alone (aged 47–79: 10-year absolute risk >2.5%), personalized screening of women age 35–79 at the same risk threshold would result in 24% fewer women being eligible for screening at a cost of 14% fewer screen-detectable cases.

Personalized screening approach could improve the efficiency of screening programs. This has potential implications on informing public health policy on cancer screening

That is, by performing SNP analysis and ten establishing a threshold one can bifurcate the groups. One could also select groups in some graded multi-sector grouping as well.

The SNPs chose are shown in a modified form below. Many are on the same gene segment. There were a total of 31 SNPs as of the date of the paper where the odds ration exceeded 1.0.

dbSNP No.	Locus/gene	Risk allele frequency	Odds Ratio per allele
rs12621278	2q31/ITGA6	0.940	1.300
rs721048	2p15	0.190	1.150
rs1465618	2p21/THADA	0.230	1.080
rs2660753	3p12	0.110	1.180
rs10934853	3q21.3	0.280	1.120
rs7679673	4q24 /TET2	0.550	1.090
rs17021918	4q22/PDLIM5	0.660	1.100
rs12500426	4q22/PDLIM6	0.460	1.080
rs9364554	6q25	0.290	1.170
rs6465657	7q21	0.460	1.120
rs10486567	7p15 /JAZF1	0.770	1.120
rs2928679	8p21	0.420	1.050
rs1512268	NKX3.1	0.450	1.180
rs620861	8q24	0.610	1.280
rs10086908	8q24	0.700	1.250
rs445114	8q24	0.640	1.140
rs16902094	8q24	0.150	1.210
rs6983267	8q24	0.500	1.260
rs16901979	8q24	0.030	2.100
rs4962416	10q26 /CTBP2	0.270	1.170
rs10993994	10q11/MSMB	0.240	1.250
rs7127900	11p15	0.200	1.220
rs7931342	11q13	0.510	1.160
rs4430796	17q12 /HNF1B	0.490	1.240
rs11649743	HNF1B	0.800	1.280
rs1859962	17q24.3	0.460	1.240
rs2735839	19q13/KLK2,KLK3	0.850	1.200
rs8102476	19q13.2	0.540	1.120
rs5759167	22q13	0.530	1.160
rs5945619	Xp11	0.280	1.120

The procedure here is an interesting first step in the genetic testing of potential cancer patients. The process however will most likely require significant refinements. The process however will most likely require significant refinements.

Thus we can ask the questions as follows:

1. Which SNPs, say the set of some n of them, provides the best set to minimize mortality and minimize the number requiring testing?

2. Can there be some clustering of SNPs such that there are disjoint classes of individuals which get assigned to risk groups. Those in the highest receiving the most significant attention and those in the lowest receiving minimal?

3. Are the SNPs such that they are independent predictors or are there environmental or other exogenous factors which can effect SNPs alone?

4. What is the relationship between SNPs and the pathways known as part of PCa development?

5. Are there temporal changes in SNPs and is there some relationship between these temporal changes? Namely are there causal SNP changes?

6. What are the causes of the SNPs?

7. Knowing the SNPs and those with PCa, what can be determined regarding the dynamics of PCa development?

8. What is the relationship between SNPs and the prostate cancer stem cell? Does the CSC have different expressions?

There are many more questions that arise from this work.

2.1 PROSTATE

In a recent 2008 NEJM article by Zheng et al they state:

"Multiple SNPs in each of the five regions were associated with prostate cancer in single SNP analysis. When the most significant SNP from each of the five regions was selected and included in a multivariate analysis, each SNP remained significant after adjustment for other SNPs and family history. Together, the five SNPs and family history were estimated to account for 46% of the cases of prostate cancer in the Swedish men we studied. The five SNPs plus family history had a cumulative association with prostate cancer ...

In men who had any five or more of these factors associated with prostate cancer, the odds ratio for prostate cancer was 9.46 …, as compared with men without any of the factors. The cumulative effect of these variants and family history was independent of serum levels of prostate-specific antigen at diagnosis...

SNPs in five chromosomal regions plus a family history of prostate cancer have a cumulative and significant association with prostate cancer."

Azzato et al have examined C1QA in breast cancer and they discuss it broadly based presence. They state:

Complement is involved in the primary defence against intravascular microorganisms and has been reported to be involved in the clearance of tumour…. Recently, we have reported an association between expression of C1QA and prognosis in oestrogen receptor (ER)-negative breast cancer… in more than one cohort.

We found that ER-negative tumours with overexpression of gene C1QA were associated with a better prognosis. The C1QA gene, located on chromosome 1p36.12, encodes for one of the components of the C1q complex. There are seven single nucleotide polymorphisms (SNPs) catalogued for C1QA on the NCBI database, of which there is only one common SNP (minor allele frequency 45%) located in an exon rs172378 is a synonymous SNP characterised by a G for A substitution at position 361 (A361G).

Thus we have another element from the immune system. It is part of the complement system, not the adaptive part and thus has primitive roots.

As Evans has noted:

However, few studies have linked SNPs with progression or prognosis of CaP. The group of Dr. William Catalona has previously reported such an association in the Journal of Urology and they continued that theme in this presentation. They investigated six newly identified genetic susceptibility variants along chromosomes 3, 5, 8, 11, and 19 to compare the pathologic tumor features between carriers and non-carriers of these newly described genetic variants. A total of 938 Caucasian men treated by radical prostatectomy between 2002 and 2008 were genotyped for CaP genetic risk variants.

The genotypes for SNPs along chromosomes 3q21 (rs10934853), 5p15 (rs401681), 8q24.21 (rs16902094[G], rs445114 [T]), 11q13 (rs11228565[A]), and 19q13.2 (rs8102476[C]) were determined for all patients, and a dominant best-fit genetic model was used to define carrier status. Pathologic tumor features were compared between carriers and noncarriers of the variants.

As Helfand et al state:

Recent studies have identified genetic variants associated with increased serum prostate specific antigen concentrations and prostate cancer risk, raising the possibility of diagnostic bias. By correcting for the effects of these variants on prostate specific antigen, it may be possible to create a personalized prostate specific antigen cutoff to more accurately identify individuals for whom biopsy is recommended. Therefore, we determined how many men would continue to meet

common biopsy criteria after genetic correction of their measured prostate specific antigen concentrations.

The genotypes of 4 single nucleotide polymorphisms previously associated with serum prostate specific antigen levels (rs2736098, rs10788160, rs11067228 and rs17632542) were determined in 964 healthy Caucasian volunteers without prostate cancer. Genetic correction of prostate specific antigen was performed by dividing an individual's prostate specific antigen value by his combined genetic risk. Analyses were used to compare the percentage of men who would meet commonly used biopsy thresholds (2.5 ng/ml or greater, or 4.0 ng/ml or greater) before and after genetic correction.

Genetic correction of serum prostate specific antigen results was associated with a significantly decreased percentage of men meeting biopsy thresholds. Genetic correction could lead to a 15% or 20% relative reduction in the total number of biopsies using a biopsy threshold of 2.5 ng/ml or greater, or 4.0 ng/ml or greater, respectively. In addition, genetic correction could result in an 18% to 22% reduction in the number of potentially unnecessary biopsies and a 3% decrease in potentially delayed diagnoses.

Our results suggest that 4 single nucleotide polymorphisms can be used to adjust a man's measured prostate specific antigen concentration and potentially delay or prevent unnecessary prostate biopsies in Caucasian men.

In a paper by as reported by Medscape they reported[5]:

In this study, 5 single-nucleotide polymorphisms (SNPs) were validated as being significantly associated with prostate cancer–specific mortality ($P \leq .05$). The 5 SNPs were located, one each, in the following 5 genes:

LEPR, the strongest marker associated with prostate cancer mortality in the study, is a cytokine receptor that is highly expressed in normal and malignant prostate tissue. The binding of leptin to its receptor leads to several downstream effects that may affect prostate carcinogenesis, including stimulation of tissue growth, inflammation, angiogenesis, and bone mass regulation. The latter effect, note the study authors, makes LEPR an interesting candidate for disease progression because the primary metastatic site for prostate cancer is the bone and bony metastases are predictive of fatal prostate cancer;

CRY1, the cryptochrome 1 gene, is in the circadian rhythm pathway, and circadian clock genes regulate androgen levels, which are known to affect prostate cancer progression and may also function as tumor suppressors through regulation of cell proliferation, apoptosis, and response to DNA damage;

RNASEL is associated with hereditary prostate cancer and is associated with apoptosis, inflammation, and cell proliferation and adhesion;

---

[5] http://www.medscape.com/viewarticle/748318_print

IL4 plays a role in cancer via activation of the Stat6 transcription factor; and

ARVCF is a member of the p120 catenin family of proteins, and increased expression has been shown to disrupt cell adhesion, which may facilitate cancer progression.

Patients with 4 to 5 at-risk genotypes had a 50% higher risk for prostate cancer–specific mortality than patients who had only 2 or fewer of these genotypes. After adjusting for clinicopathological factors know to affect prognosis, the risk for mortality increased with the number of at-risk genotypes (P for trend = .001).

As Lin et al in the paper report:

Five SNPs were validated (P ≤ 0.05) as being significantly associated with PCSM, one each in the LEPR, CRY1, RNASEL, IL4, and ARVCF genes. Compared with patients with 0 to 2 of the at-risk genotypes those with 4 to 5 at-risk genotypes had a 50% (95% CI, 1.2–1.9) higher risk of PCSM and risk increased with the number of at-risk genotypes carried (Ptrend = 0.001), adjusting for clinicopathologic factors known to influence prognosis.

Five genetic markers were validated to be associated with lethal prostate cancer. This is the first population-based study to show that germline genetic variants provide prognostic information for prostate cancer-specific survival. The clinical utility of this five-SNP panel to stratify patients at higher risk for adverse outcomes should be evaluated.

In Biotechniques they report on the Eeles work as follows[6]:

In genome-wide association studies, the bigger, the better—this week, a mega-consortium has published the results from the biggest cancer genotyping effort to date.

Thirteen papers, based on genetic data from 200,000 individuals, are being published today (March 27) in Nature Genetics, Human Molecular Genetics, Nature Communications, PLoS Genetics, and The American Journal of Human Genetics. The papers report 74 new genetic variants, or single nucleotide polymorphisms (SNPs), associated with three hormone-related cancers: breast, prostate, and ovarian cancer….

Finally, in a study of 25,000 prostate cancer cases compared to 25,000 controls, Rosalind A. Eeles at the The Institute of Cancer Research and colleagues identified 23 new prostate cancer SNPs of significance, 16 of which are associated with aggressive disease (5). Seventy-eight prostate cancer susceptibility SNPs have now been identified, and these genetic variants explain 36 percent of the familial risk for the disease.

The total known SNPs, including those newly identified, explain roughly 35-50 percent of the heritability of each disease, as noted above, but that leaves over 50 percent of the heritability unexplained. This is likely due to additional SNPs that haven't yet been identified, including rare variants, said Easton. For example, in breast cancer 76 SNPs have now been identified, but

[6] http://www.biotechniques.com/news/SNPping-Away-at-Cancer/biotechniques-341608.html?service=print

"there are probably several thousand of these, at least," he said. Discovering those unknown SNPs will likely require both more analysis of the current data and larger genome-wide association studies in the future, he added.

From the paper by Eeles et al we have the following putative SNPs:

SNP Marker	***Chromosome***	***Allele***	***Candidate Gene***
rs11135910	8	25948059	*EBF2*
rs11568818	11	101906871	*MMP7*
rs11650494	17	44700185	*HOXB13*, *PRAC*, *SPOP*, *ZNF652*
rs11902236	2	10035319	*TAF1B*, *GRHL1*
rs12155172	7	20961016	*SP8*
rs1218582	1	153100807	*KCNN3*
rs1270884	12	113169954	*TBX5*
rs1894292	4	74568022	*AFM*, *RASSF6*
rs1933488	6	153482772	*RSG17*
rs2273669	6	109391882	*ARMC2*, *SESN1*
rs2405942	X	9774135	*SHROOM2*
rs2427345	20	60449006	*GATAS*, *CABLES2*
rs3096702	6	32300309	*NOTCH4*
rs3771570	2	242031537	*FARP2*
rs3850699	10	104404211	*TRIM8*
rs4245739	1	202785465	*MDM4*
rs6062509	20	61833007	*ZGPAT*
rs684232	17	565715	*VPS53*, *FAM57A*
rs6869841	5	172872032	*BOD1* (*FAM44B*)
rs7141529	14	68196497	*RAD51B*
rs7241993	18	74874961	*SALL3*
rs7611694	3	114758314	*SIDT1*
rs8008270	14	52442080	*FERMT2*

Penney et al show several specific SNPs building upon earlier work of Eeles et al. delineating lowered risk:

Recent genome-wide association studies (GWAS) identified numerous genetic regions associated with prostate cancer risk, including the locus on chromosome 19 containing KLK3. The minor allele of the single-nucleotide polymorphism (SNP) rs2735839 in this region was associated with a 17–46% decreased risk of prostate cancer; Eeles et al. initially reported these results in a multistage design in which stage 1 of the study specifically excluded controls with PSA levels .0.5 ng/ml.

The goal of this sampling strategy was to decrease the possibility of including controls with occult prostate cancer, but this design raised the question of whether this genetic variant was truly a prostate cancer risk factor or merely associated with the likelihood of diagnosis by virtue of its link with PSA levels.

Several groups have observed an association between rs2735839 and PSA levels. Ahn et al. genotyped tag SNPs across KLK3 and observed no association with prostate cancer risk in several populations; they only saw an association of rs2735839 when they artificially restricted controls to those with PSA levels ,0.5 ng/ml.

From the work of Jin et al we have the following list:

Chromosome Number	***SNP***	***Region***	***Position***	***Gene***	***Allele***	***Risk Allele***	***RAF***
7	rs10486567	7p15	27,943,088	JAZF1	C/T	C	0.778
11	rs10896449	11q13	68,751,243		G/A	G	0.549
10	rs10993994	10q11	51,219,502	MSMB	C/T	T	0.427
17	rs11649743	17q12 (region 2)	33,149,092	HNF1B	C/T	C	0.838
8	rs1447295	8q24 (region 1)	128,554,220		C/A	A	0.163
9	rs1571801	9p13	123,467,194	DAB2IP	G/T	T	0.278
8	rs16901979	8q24 (region 2)	128,194,098		C/A	A	0.062
17	rs1859962	17q24	66,620,348		G/T	G	0.526
3	rs2660753	3p12	87,193,364		C/T	T	0.110
19	rs2735839	19q13	56,056,435	KLK2/KLK3	G/A	G	0.869
17	rs4430796	17q12 (region 1)	33,172,153	HNF1B	T/C	T	0.590
10	rs4962416	10q26	126,686,862	CTBP2	A/G	G	0.267
X	rs5945619	Xp11	51,074,708	NUDT10/NUDT11	A/G	G	0.428
7	rs6465657	7q21	97,654,263	LMTK2	T/C	C	0.491
8	rs6983267	8q24 (region 3)	128,482,487		G/T	G	0.561
2	rs721048	2p15	62,985,235	EHBP1	G/A	A	0.178
6	rs9364554	6q25	106,280,983	SLC22A3	C/T	T	0.309

The totality of all of these works lets us see that there is a growing body of evidence that SNPs are related to PCa. The question is why and what do they do. We understand many of the PCa pathways and many of these identified SNPs relate to genes not in those pathways, at least not at this time. The challenge will be to identify the causative pathways elements, not just speculate on SNPs.

One clear observation from this very small sample of studies is the plethora of putative SNPs related to PCa. There is a nexus with many genes, but these genes are not currently key to PCa pathway problems. As we have stated before:

1. Models for PCa genetic system breakdown have identified genes and often the genes related in SNP studies are not them.

2. The causative nature of the identified SNPs is missing.

3. The existence and cause of the SNPs is also not identified. One suspects that many of the SNPs are germline in nature and thus one wonders what the cause and effect relationships are.

2.2 MELANOMA

We now want to consider melanoma and the SNP issue. This gives us an epithelial cancer.

From Barrett et al we have:

We report a genome-wide association study for melanoma that was conducted by the GenoMEL Consortium. Our discovery phase included 2,981 individuals with melanoma and 1,982 study-specific control individuals of European ancestry, as well as an additional 6,426 control subjects from French or British populations, all of whom were genotyped for 317,000 or 610,000 single-nucleotide polymorphisms (SNPs).

Our analysis replicated previously known melanoma susceptibility loci. Seven new regions with at least one SNP with $P < 10^{-5}$ *and further local imputed or genotyped support were selected for replication using two other genome-wide studies (from Australia and Texas, USA). Additional replication came from case-control series from the UK and The Netherlands.*

Variants at three of the seven loci replicated at $P < 10^{-3}$*:*

an SNP in <u>ATM</u> (rs1801516, overall $P = 3.4 \times 10^{-9}$***),***

an SNP in <u>MX2</u> (rs45430, $P = 2.9 \times 10^{-9}$***) and***

an SNP adjacent to <u>CASP8</u> (rs13016963, $P = 8.6 \times 10^{-10}$***).***

A fourth locus near CCND1 remains of potential interest, showing suggestive but inconclusive evidence of replication (rs1485993, overall $P = 4.6 \times 10^{-7}$ *under a fixed-effects model and* $P = 1.2 \times 10^{-3}$ *under a random-effects model).*

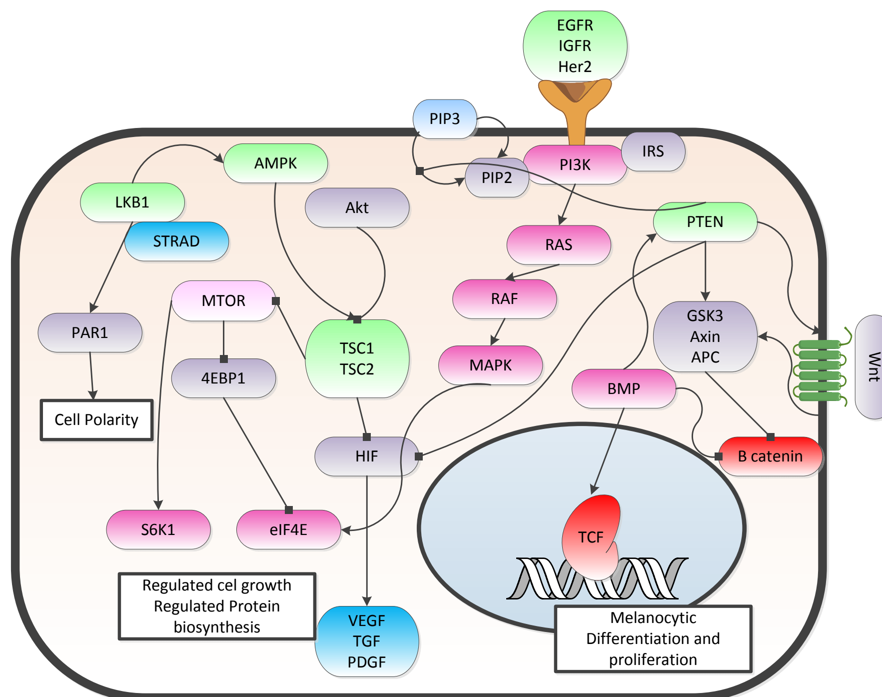
In the article by Liu et al the authors state:

Genome-wide association studies (GWASs) have mainly focused on top significant single nucleotide polymorphisms (SNPs), most of which did not have clear biological functions but were just surrogates for unknown causal variants.

Studying SNPs with modest association and putative functions in biologically plausible pathways has become one complementary approach to GWASs.

To unravel the key roles of mitogen-activated protein kinase (MAPK) pathways in cutaneous melanoma (CM) risk, we re-evaluated the associations between 47 818 SNPs in 280 MAPK genes and CM risk using our published GWAS dataset with 1804 CM cases and 1026 controls. We initially found 105 SNPs with $P \leq 0.001$*, more than expected by chance, 26 of which were predicted to be putatively functional SNPs.*

From Bauer and Straikis we show the MAPK pathway in melanoma below.



From: Bauer and Stratakis

Bauer and Straikis state:

Investigation of one of the signalling pathways, the mitogen activated protein kinase (MAPK) ERK 1/2 pathway, typically inhibited by PKA in many cells, has recently been reported. In this report, the lymphocytes from CNC patients with known PRKAR1A mutations showed altered PKA activity and increased ERK 1/2 phosphorylation. Cell metabolism and cell proliferation studies suggested that altered PKA activity is associated with reversal of PKA mediated inhibition of the MAPK pathway resulting in increased cell proliferation

Now Liu et al continue:

*The risk associations with 16 SNPs around DUSP14 (rs1051849) and a previous reported melanoma locus MAFF/PLA2G6 (proxy SNP rs4608623) were replicated in the GenoMEL dataset (*P *< 0.01) but failed in the Australian dataset. Meta-analysis showed that rs1051849 in the 3′ untranslated regions of DUSP14 was associated with a reduced risk of melanoma (odds ratio = 0.89, 95% confidence interval: 0.82–0.96,* P *= 0.003, false discovery rate = 0.056). Further genotype–phenotype correlation analysis using the 90 HapMap lymphoblastoid cell lines from Caucasians showed significant correlations between two SNPs (rs1051849 and rs4608623) and messenger RNA expression levels of DUSP14 and MAFF (*P *= 0.025 and* P *= 0.010, respectively).*

Gene-based tests also revealed significant SNPs were over-represented in MAFF, PLA2G6, DUSP14 and other 16 genes. Our results suggest that functional SNPs in MAPK pathways may contribute to CM risk. Further studies are warranted to validate our findings.

This paper is somewhat different from others in that they associate SNPs wi8th specific pathway operations and changes. The question is what a SNP does that results in a breakdown of a homeostatic pathway and result in a melanoma. These details will be critical.

3 SNP-SNP INTERACTIONS

The previous discussions were focused on single SNPs. In a recent paper by H. Lin the authors discuss SNP-SNP interactions. The paper starts out making a strong statement regarding the current state of understanding of prostate cancer prognosis. Specifically:

When using the existing features, approximately 20% of these low-risk prostate cancer patients died due to conservative treatment. Thus, there is an urgent need for identifying biomarkers in order to improve prediction accuracy of prostate cancer aggressiveness.

Namely the term "low risk" is a misnomer. With a 20% failure rate this provides many men with a deadly prognosis based upon total ignorance. Their hopes is to improve that situation. Their approach is to focus on SNP pairing and specifically how that relates to angiogenesis. They state[7]:

Angiogenesis is a biological process that involves the division and migration of endothelial cells, resulting in microvasculature formation. The formation of blood vessels is important for organ development during embryogenesis and continues to contribute to organ growth after birth. During adulthood, most blood vessels remain quiescent and angiogenesis is limited to the cycling ovary and in the placenta during pregnancy.

Nonetheless, endothelial cells maintain their ability to divide rapidly into blood vessels in response to physiological stimuli, such as hypoxia, and angiogenesis is reactivated during wound healing and repair.

The process of postnatal angiogenesis is regulated by a continuous interplay (that establishes a balance) of stimulators such as;

1. *vascular endothelial growth factor (VEGF),*
2. *basic fibroblast growth factor (bFGF),*
3. *epidermal growth factor (EGF),*
4. *interleukins (ILs),*
5. *transforming growth factor beta (TGF-b),*
6. *tumor necrosis factor alpha (TNF-a),*
7. *platelet derived growth factor (PDGF), and*
8. *matrix metalloproteinases (MMPs) and*
9. *inhibitors such as endostatin, platelet factor-4, tumastin, thrombospondin- 1, plasminogen activator inhibitor-1 and angiostatin .*

However, in many disorders including prostate cancer, the balance between stimulators and inhibitors is tilted to favor stimulators, resulting in an ''angiogenic switch''

[7] Note: We have modified the formatting to present the results in a more direct manner. We have not made any changes to the words.

Their overall conclusions are:

Angiogenesis has been shown to be associated with prostate cancer development. The majority of prostate cancer studies focused on individual single nucleotide polymorphisms (SNPs) while SNP-SNP interactions are suggested having a great impact on unveiling the underlying mechanism of complex disease.

Using 1,151 prostate cancer patients in the Cancer Genetic Markers of Susceptibility (CGEMS) dataset, 2,651 SNPs in the angiogenesis genes associated with prostate cancer aggressiveness were evaluated. SNP-SNP interactions were primarily assessed using the two-stage Random Forests plus Multivariate Adaptive Regression Splines (TRM) approach in the CGEMS group, and were then re-evaluated in the Moffitt group with 1,040 patients.

For the identified gene pairs, cross-evaluation was applied to evaluate SNP interactions in both study groups. Five SNP-SNP interactions in three gene pairs (MMP16+ ROBO1, MMP16+ CSF1, and MMP16+ EGFR) were identified to be associated with aggressive prostate cancer in both groups.

Three pairs of SNPs (rs1477908+ rs1387665, rs1467251+ rs7625555, and rs1824717+ rs7625555) were in MMP16 and ROBO1, one pair (rs2176771+ rs333970) in MMP16 and CSF1, and one pair (rs1401862+ rs6964705) in MMP16 and EGFR.

The results suggest that MMP16 may play an important role in prostate cancer aggressiveness. By integrating our novel findings and available biomedical literature, a hypothetical gene interaction network was proposed. This network demonstrates that our identified SNP-SNP interactions are biologically relevant and shows that EGFR may be the hub for the interactions.

The findings provide valuable information to identify genotype combinations at risk of developing aggressive prostate cancer and improve understanding on the genetic etiology of angiogenesis associated with prostate cancer aggressiveness.

We summarize these results in the matrix Table below.

	MMP16	ROBO1	CSF1	EGFR
MMP16		rs1477908+ rs1387665, rs1467251+ rs7625555, and rs1824717+ rs7625555	rs2176771+ rs333970	rs1401862+ rs6964705
ROBO1	rs1477908+ rs1387665, rs1467251+ rs7625555, and			

	rs1824717+ rs7625555			
CSF1	rs2176771+ rs333970			
EGFR	rs1401862+ rs6964705			

3.1 SNP-SNP Analysis

Lin et al conclude:

Our findings identified five SNP-SNP interactions in the angiogenesis genes associated with prostate cancer aggressiveness in the CGEMS group using the novel TRM approach.

Five highly significant SNP-SNP interactions (p-value = 261025 to 661024) with a medium to large effect size were successfully detected even with a relatively small sample size of approximately 1,000.

The odds ratios of these SNP interactions were categorized from a medium (OR>1.5) to large effect size (OR>2). The clinical impact of the SNP-SNP interactions may be larger than that for individual SNPs identified in GWA studies.

The prediction power of cancer risk for the SNPs identified in GWA studies is limited with the median per-allele OR of 1.22 based on a recent review. Our identified gene-gene interactions may be biologically relevant based on the network analysis.

The interactions of the five gene pairs:

1. *MMP16+ ROBO1,*
2. *MMP16+ CSF1,*
3. *MMP16+ EGFR,*
4. *CSF1+ FBLN5, and*
5. *CSF1+ HSPG2*

were demonstrated using cross-evaluation in the CGEMS and Moffitt groups. Particularly, the former three gene pairs had at least one SNPSNP interaction with a similar interaction pattern in the two study groups.

Gene	Location
CSF1 ***colony stimulating factor 1***	*1p13.3*
HSPG2 ***heparan sulfate proteoglycan 2***	*1p36*
ROBO1 ***roundabout, axon guidance receptor, homolog 1***	*3p12*
EGFR ***epidermal growth factor receptor***	*7p12*
MMP-16 ***matrix metallopeptidase 16 (membrane-inserted)***	*8q21.3*
FBLN5 ***fibulin 5***	*14q32.1*

We now show the related SNPs and their locations. Note than there are none on Chromosome 14.

SNP RS#[8]	Chromosome	Phys_loc (bp)	Gen_loc (cM)
rs333970	1	110178380	140.418328
rs1387665	3	79512501	110.02294
rs7625555	3	78759095	109.739119
rs6964705	7	54983846	77.813885
rs1401862	8	89273954	113.321995
rs1467251	8	89341178	113.388945
rs1477908	8	89142200	113.190777
rs1824717	8	89145095	113.193661
rs2176771	8	89276381	113.324412

In the Table below we present substantial details on each of these key genes.

Gene	Location	Function[9]
MMP-16 matrix metallopeptidase 16 (membrane-inserted)	*8q21.3*	*Proteins of the matrix metalloproteinase (MMP) family are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. Most MMP's are secreted as inactive proproteins which are activated when cleaved by extracellular proteinases. The encoded protein activates MMP2 by cleavage. This gene was once referred to as MT-MMP2, but was renamed as MT-MMP3 or MMP16.*

[8] http://integrin.ucd.ie/cgi-bin/rs2cm.cgi Note that one can enter a list of SNPs and obtain locations. Also note one can access details from *http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1401862*.

[9] http://www.ncbi.nlm.nih.gov/gene/4325

Gene	Location	Function[9]
ROBO1 roundabout, axon guidance receptor, homolog 1	*3p12*	*Bilateral symmetric nervous systems have special midline structures that establish a partition between the two mirror image halves. Some axons project toward and across the midline in response to long-range chemoattractants emanating from the midline. The product of this gene is a member of the immunoglobulin gene superfamily and encodes an integral membrane protein that functions in axon guidance and neuronal precursor cell migration. This receptor is activated by SLIT-family proteins, resulting in a repulsive effect on glioma cell guidance in the developing brain. A related gene is located at an adjacent region on chromosome 3. Multiple transcript variants encoding different isoforms have been found for this gene.*
CSF1 colony stimulating factor 1	*1p13.3*	*The protein encoded by this gene is a cytokine that controls the production, differentiation, and function of macrophages. The active form of the protein is found extracellularly as a disulfide-linked homodimer, and is thought to be produced by proteolytic cleavage of membrane-bound precursors. The encoded protein may be involved in development of the placenta. Alternate splicing results in multiple transcript variants.*
EGFR epidermal growth factor receptor	*7p12*	*The protein encoded by this gene is a transmembrane glycoprotein that is a member of the protein kinase superfamily. This protein is a receptor for members of the* ***epidermal growth factor family. EGFR is a cell surface protein that binds to epidermal growth factor.*** *Binding of the protein to a ligand induces receptor dimerization and tyrosine autophosphorylation and leads to cell proliferation. Mutations in this gene are associated with lung cancer. Multiple alternatively spliced transcript variants that encode different protein isoforms have been found for this gene.*
FBLN5 fibulin 5	*14q32.1*	*The protein encoded by this gene is a secreted, extracellular matrix protein containing an Arg-Gly-Asp (RGD) motif and calcium-binding EGF-like domains. It promotes adhesion of endothelial cells through interaction of integrins and the RGD motif. It is prominently expressed in developing arteries but less so in adult vessels. However, its expression is reinduced in balloon-injured vessels and atherosclerotic lesions, notably in intimal vascular smooth muscle cells and endothelial cells. Therefore, the* ***protein encoded by this gene may play a role in vascular development and remodeling.*** *Defects in this gene are a cause of autosomal dominant cutis laxa, autosomal recessive cutis laxa type I (CL type I), and age-related macular degeneration type 3 (ARMD3).*

Gene	*Location*	*Function*[9]
HSPG2 heparan sulfate proteoglycan 2	*1p36*	*This gene encodes the perlecan protein, which consists of a core protein to which three long chains of glycosaminoglycans (heparan sulfate or chondroitin sulfate) are attached. The perlecan protein is a large multidomain proteoglycan that binds to and cross-links many extracellular matrix components and cell-surface molecules. It has been shown that this protein interacts with laminin, prolargin, collagen type IV, FGFBP1, FBLN2, FGF7 and Transthyretin, etc. and plays essential roles in multiple biological activities. Perlecan is a key component of the vascular extracellular matrix, where it helps to maintain the endothelial barrier function. It is a potent inhibitor of smooth muscle cell proliferation and is thus thought to help maintain vascular homeostasis.* ***It can also promote growth factor (e.g., FGF2) activity and thus stimulate endothelial growth and re-generation. It is a major component of basement membranes, where it is involved in the stabilization of other molecules as well as being involved with glomerular permeability to macromolecules and cell adhesion.*** *Mutations in this gene cause Schwartz-Jampel syndrome type 1, Silverman-Handmaker type of dyssegmental dysplasia, and Tardive dyskinesia*

Furthermore Lin et al state:

Angiogenesis has been shown to be associated with prostate cancer development. The majority of prostate cancer studies focused on individual single nucleotide polymorphisms (SNPs) while SNP-SNP interactions are suggested having a great impact on unveiling the underlying mechanism of complex disease.

Using 1,151 prostate cancer patients in the Cancer Genetic Markers of Susceptibility (CGEMS) dataset, 2,651 SNPs in the angiogenesis genes associated with prostate cancer aggressiveness were evaluated. SNP-SNP interactions were primarily assessed using the two-stage Random Forests plus Multivariate Adaptive Regression Splines (TRM) approach in the CGEMS group, and were then re-evaluated in the Moffitt group with 1,040 patients. For the identified gene pairs, cross-evaluation was applied to evaluate SNP interactions in both study groups.

Five SNP-SNP interactions

in three gene pairs

(MMP16+ ROBO1, MMP16+ CSF1, and MMP16+ EGFR) were identified to be associated with aggressive prostate cancer in both groups.

I. Three pairs of SNPs

1. *rs1477908+ rs1387665,*
2. *rs1467251+ rs7625555, and*
3. *rs1824717+ rs7625555*

were in ***MMP16 and ROBO1****,*

II. one pair

4. rs2176771+ rs333970
*in **MMP16 and CSF1**, and*

III. one pair
5. rs1401862+ rs6964705
*in **MMP16 and EGFR**.*

The results suggest that MMP16 may play an important role in prostate cancer aggressiveness. By integrating our novel findings and available biomedical literature, a hypothetical gene interaction network was proposed.

*This network demonstrates that our identified SNP-SNP interactions are biologically relevant and **shows that EGFR may be the hub for the interactions**. The findings provide valuable information to identify genotype combinations at risk of developing aggressive prostate cancer and improve understanding on the genetic etiology of angiogenesis associated with prostate cancer aggressiveness.*

3.2 Specific Genes

We now will examine each of the six paired genes to demonstrate functionality and pathway impact as may be appropriate.

MMP-14	ROBO1	CSF1	EGFR	FBLN5	HSPG2
• Matrix metalloproteinase is a gene whose product breaks down celular elements and is found in many metastatic cancers.	• A gene whose product is in the immunoglobulin family of genes.	• The gene codes for a cytokine, one of the elements in the immune system.	• EGFR initiates cell motility and movement. Movement of the cell is one of the first steps in a cell becoming metastatic.	• A gene whose product is in the ECM and initiates the process of angiogenesis	• Also known as perlecan and it is found in the ECM and is associated with adhesion. It can also stimulate cell growth. Overexpression may cause vascularization of cell clusters.

We now proceed to detail each of these genes.

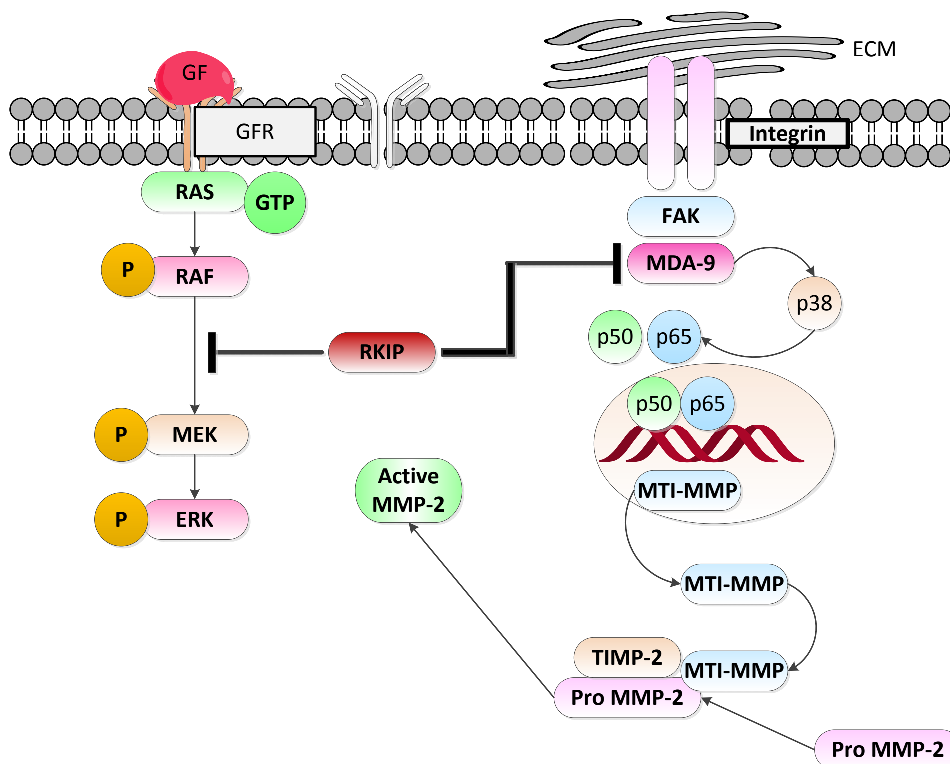
3.2.1 MMP

MMPs are matrix metalloproteinases. They work in the ECM. We refer the reader to our paper on the ECM for substantial discussion there.

As Hagase states:

Matrix metalloproteinases (MMPs), also called matrixins, function in the extracellular environment of cells and degrade both matrix and non-matrix proteins. ***They play central roles in morphogenesis, wound healing, tissue repair and remodeling in response to injury****, e.g. after myocardial infarction, and in progression of diseases such as atheroma, arthritis, cancer and chronic tissue ulcers. They are multi-domain proteins and their activities are regulated by tissue inhibitors of metalloproteinases (TIMPs).*

… members of the MMP family and discusses their domain structure and function, proenyme activation, the mechanism of inhibition by TIMPs and their significance in physiology and pathology….MMP-4, MMP-5, MMP-6 and MMP-22 are missing in the list since they were shown to be identical to other members.



From Delassus et al:

AP-2alpha, interleukin-4 (IL-4), E-cadherin, fibulin 1D, p16(INK4alpha), PTEN, RKIP, and S100A4 are determinants (suppressors, except for S100A4) of cancer cell invasiveness and other traits of cancer progression, which are located upstream of matrix metalloproteinases (MMPs) in cell signaling pathways.

We will refer to them as upstream cancer-progression determinants (UCPDs, for brevity). MMP-1, MMP-2, MMP-9, MMP-11, MMP-13, MMP-14, MMP-16, and MMP-19 are enhancers of cancer cell invasiveness and other traits of cancer progression, in MDA-MB-231 breast cancer cells.

We are interested in pathway links from UCPDs to gene expression of cancer cell MMPs in MDA-MB-231 cells. To test models about these links, wild-type copies of UCPDs were transiently overexpressed and then MMP mRNAs were measured by reverse transcription real-time PCR. The present results show that each of eight UCPDs is linked to the gene expression of a unique set of MMPs. This indicates that the effects are sequence-specific and that each UCPD reaches these MMP expressions through different sets of signaling pathways.

We have detected 20 new pathway links, 11 are downregulatory and nine are upregulatory; 15 are new links in any cell, and five are new links in breast cancer. In seven links, three cancer-progression suppressing UCPDs unexpectedly enhance the gene expression of five cancer-progression promoting MMPs.

3.2.2 ROBO1

As Dickinson and Duncan state:

The secreted SLIT glycoproteins and their Roundabout (ROBO) receptors were originally identified as important axon guidance molecules. They function as a repulsive cue with an evolutionarily conserved role in preventing axons from migrating to inappropriate locations during the assembly of the nervous system. In addition the SLIT-ROBO interaction is involved in the regulation of cell migration, cell death and angiogenesis and, as such, has a pivotal role during the development of other tissues such as the lung, kidney, liver and breast.

The cellular functions that the SLIT/ROBO pathway controls during tissue morphogenesis are processes that are dysregulated during cancer development. Therefore inactivation of certain SLITs and ROBOs is associated with advanced tumour formation and progression in disparate tissues.

As regards to the pathway issues SLIT as follows as per Dickinson and Duncan:

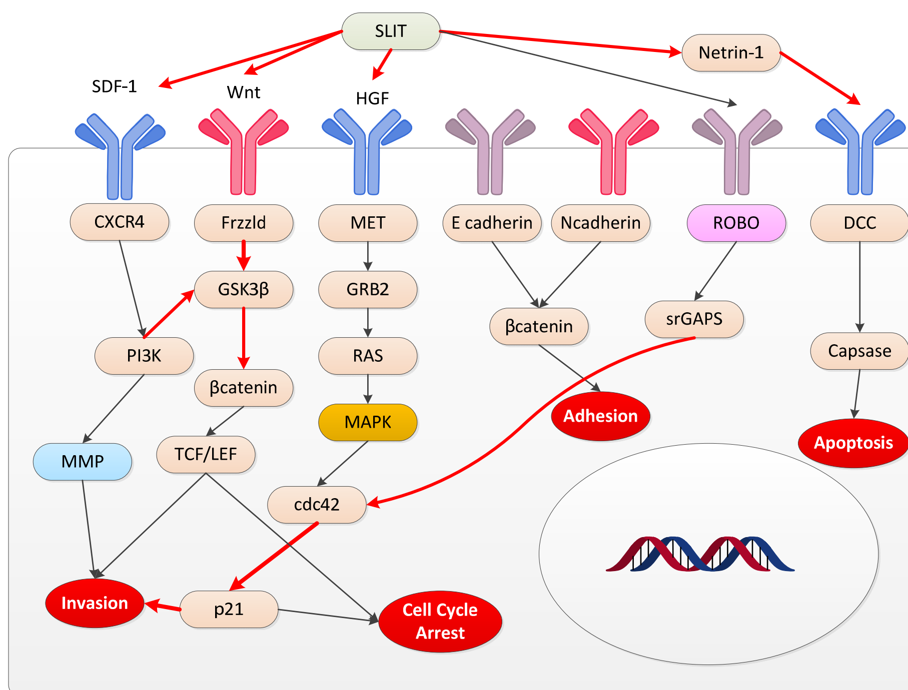
LIT can inhibit invasion and promote a cell cycle arrest by blocking Wnt, HGF and SDF-1 signalling. The SLIT-ROBO can also prevent invasion and stimulate a cell cycle arrest directly

by negatively regulating cdc42 activity. SLIT binding to ROBO also relieves inhibition of DCC by Netrin-1.

This allows the activation of pro-apoptotic pathways through Caspase 3 and 9. SLIT can also bind and sequester Netrin-1 preventing its interaction with DCC and inhibitory role in apoptosis. Depending on the particular cellular environment, the SLIT-ROBO interaction can also promote and inhibit adhesion. The SLIT-ROBO interaction promotes adhesion in breast tumour cells and during mammary gland development, possibly by enhancing the association between E-cadherin and β-catenin at cell borders.

However during formation of the heart lumen, SLIT-ROBO signalling antagonises E-cadherin/β-catenin mediated cell-cell adhesion. During neural development SLIT binding promotes an interaction between ROBO and N-cadherin. Subsequently β- catenin becomes disassociated from the complex and there is a reduction in cadherin mediated cell-cell adhesion. ***Black arrows represent promoting an activity while red arrows depict inhibiting an action.***

We depict below the relationship of SLIT to ROBO and then ROBO to CDC42 to p21 and then Invasion. We depict that below:



From: Dickinson and Duncan

This Figure for the pathways shows most of the key elements proposed. It shows MDM, SLIT, and ROBO.

3.2.3 CSF1

CSF or the colony stimulating factor is an aggressive gene product that enhances cell growth.

As Dey et al state:

Colony-stimulating factor-1 (CSF-1) induces expression of immediate early gene, such as c-myc and c-fos and delayed early genes such as D-type cyclins (D1 and D2), whose products play essential roles in the G1 to S phase transition of the cell cycle. Little is known, however, about the cytoplasmic signal transduction pathways that connect the surface CSF-1 receptor to these genes in the nucleus.

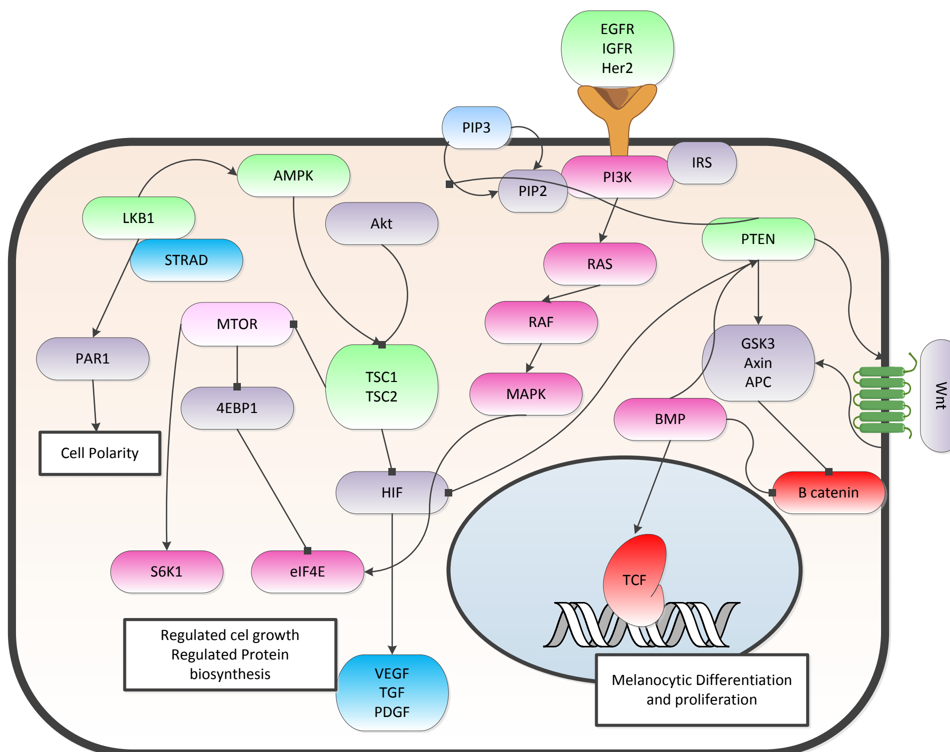
We have investigated the signaling mechanism of CSF-1-induced D2 expression. Analyses of CSF-1 receptor autophosphorylation mutants show that, although certain individual mutation has a partial inhibitory effect, only multiple combined mutations completely block induction of D2 in response to CSF-1.

We report that at least three parallel pathways, the Src pathway, the MAPK/ERK kinase (MEK)/extracellular signal-regulated kinase (ERK) pathway, and the c-myc pathway, are involved. Induction of D2 is partially inhibited in Src2/2 bone marrow-derived macrophages and by Src inhibitor PP1 and is enhanced in v-Src-overexpressing cells. Activation of myc's transactivating activity selectively induces D2 but not D1. Blockade of c-myc expression partially blocks CSF-1-induced D2 expression.

Complete inhibition of the MEK/ERK pathway causes 50% decrease of D2 expression. Finally, simultaneous inhibition of Src, MEK activation, and c-myc expression additively blocks CSF-1-induced D2 expression. This study indicates that multiple signaling pathways are involved in full induction of a single gene, and this finding may also apply broadly to other growth factor-inducible genes.

3.2.4 EGFR

The EGFR growth factor receptor is a major receptor in cells. We depict its key pathways below.



The above has been discussed in detail in our works on melanoma and prostate cancer system genomics.

3.2.5 FBLN5

From Schulterman:

Fibulin-5 (Fbln5) is a matricellular protein recently shown to regulate angiogenesis; however its effect on tumor angiogenesis and thus tumor growth is currently unknown. ...

Matricellular proteins are expressed at sites of tissue remodeling where they coordinate cell-ECM interaction. As such this unique class of proteins is well-suited to influence the TME and tumor progression. Much of our understanding about the function of matricellular proteins in tumorigenesis is a result of studies in mice engineered to lack the expression of specific matricellular proteins including SPARC, thrombospondin-1 (TSP-1), osteopontin (OPN) and fibulin-5 (Fbln5). Although it is clear that these proteins influence tumor growth…

Fibulin-5 (Fbln5) is a matricellular protein required for maturation of elastic fibers, which provide elasticity to the blood vessel wall. Therefore, Fbln5 has a direct effect on the efficiency of the vasculature. But it is its unique ability to alter cell-ECM signaling via integrin binding and the downstream affects this has on angiogenesis that make it an interesting protein for cancer research.

Fbln5 (Dance, EVEC) is a member of the fibulin family of ECM proteins which all contain a string of calcium-binding epidermal growth factor-like (cbEGF) repeats at the N-terminus followed by the defining globular COOHterminal fibulin-type module. cbEGF motifs have been shown to be important for proper protein folding and structure stabilization and act as signaling sequences for protein interaction. To date, the function for the fibulin-type module is unknown.

The high level of Fbln5 expression by endothelial cells and the aberrant vessel defect in Fbln5-/- mice indicates a critical function for Fbln5 in the vascular environment. Initial in vitro studies denoted Fbln5 as an inhibitor of angiogenesis. Treatment with recombinant Fbln5 inhibited the proliferation and invasion of murine brain microvascular endothelial cells through matrigel by antagonizing VEGF activation of the ERK1/ERK2 signaling pathway. Furthermore, Fbln5 was shown to be a target of TGF and induced the expression of the anti-angiogenic protein TSP-1. In addition, activated endothelial cells undergoing tubulogenesis downregulated expression of Fbln5.

Investigation into the effect of Fbln5 on tumor development and progression are still in the early stages and little is currently known. However, the expression of Fbln5 in human cancers has recently been examined in a small subset of tumor types including kidney, breast, lung, ovary and some gastrointestinal cancers. In this study, Fbln5 mRNA was evaluated using a cDNA microarray coated with matched normal/tumor cDNA from 68 patients with varying cancers. Fbln5 expression was altered in 44 of 68 samples and of those 44 cases; expression was down-regulated in 42 and up-regulated in only two. 63 It is important to note, however, that the samples examined in this study were derived from whole tumors. Therefore

3.2.6 HSPG2

HSPG2 is also known as perlecan.

As Maeshima states:

In the process of ECM remodeling during vessel sprouting, proteases such as urokinase plasminogen activator (uPA) and its inhibitor PAI-1, matrix metalloproteinases (MMPs) and tissue inhibitor of MMP, TIMPs, heparinases, and cathepsins play important roles (64, 65). HSPGs present in the basement membranes sequester proangiogenic growth factors such as VEGF and bFGF.

Proteinases liberate matrix-bound proangiogenic factors, thus facilitating sprouting angiogenesis. MMP-9 and MMP-2 are known to be required for the mobilization of the sequestered VEGF and thus the initiation of tumor angiogenesis…

Perlecan is a major HSPG of basement membranes and vascular and avascular ECM, involved in regulating cell growth, differentiation, cell adhesion, and the development of blood vessels, cartilage, and the nervous system.

Perlecan-null mutations result in early embryonic lethality accompanied by severe cephalic and cartilage abnormalities . Embryos that survive initially usually develop later malformations of

the cardiovascular system . In humans, two rare skeletal disorders, dyssegmental dysplasia silver-handmaker type (DSSH) and Schwartz-Jampel syndrome (SJS) are caused by mutations of genes encoding perlecan. Perlecan is considered to exert proangiogenic effects, because it binds to and protect growth factors from degradation and it interacts with adhesion molecules ….

As Datta et al state:

Perlecan is expressed in and associated with aggressive prostate cancers After identification of Perlecan as a candidate gene for the CAPB locus we sought to confirm the presence of Perlecan in primary prostate cancers. Immunohistochemical analysis for Perlecan in prostate cancer tissue microarrays with 600 patient samples demonstrated that Perlecan, a secreted proteoglycan, is present in the lumens of 54% of malignant prostate cancer glands, but not in normal glands (Figure. 1A–D, Table 1).

There was a significant increase in Perlecan levels in invasive tumors compared to either benign prostate tissue or the precancerous lesion high grade prostatic intraepithelial neoplasia (HGPIN). In particular Perlecan expression was associated with more aggressive tumors, as evidenced by their higher Gleason score

3.3 THE CLUSTERING APPROACH

The authors of the SNP-SNP work use a clustering approach of MARS and Forests. We examine them here. Unfortunately they do not present the details of their analysis and thus one cannot independently evaluate the results. This seems to be a common practice in many such studies.

We start with the analysis of Selinski and Ickstadt. We have modified it a bit[10].

Let us assume we have N SNP locations.

We have two classes of patients, those with a specific disorder and those without. We examine a specific SNP location. We assume it is the same SNP location for each set. Now define:

$$V_1 = \{P; st\ P\ are\ normal\}$$
$$V_2 = \{P; st\ P\ are\ abnormal\ or\ diseased\}$$

Now for simplicity assume we have M patients in each class.

We desire a measure of matching, namely a measure that maps the number of times the two groups differ[11]. A simple approach is to do the following:

[10] One should also examine Murphy for an encyclopedic presentation of the many types of approaches. It is not clear that any are truly preferable, their acceptance is often almost religious in acceptance. The tree and forest approach of the SNP-SNP paper is one such example.

[11] There are many measures for SNP distances. We have Muller, Selinski and Ickstadt as examples of these types.

Let SNP_k equal the SNP value on the normal patients.

Let us consider collecting data on a set of patients with cancer and those without. Let us assume for the moment that it is prostate cancer. We then select data on a collection of SNPs in both cases. The following describes the sample from M1 patients having prostate cancer. The entry is 0 if the SNP from that patient is what are a common SNP and 1 otherwise. Now we may have the following 2 example data sets, the H1 set for the cancer cases and the H0 set for the non-cancer cases. These are samples for just single SNPs. It should also be noted that we can have a measure that is 0,1,and 2, based upon no SNP, one SNP on one chromosome, and 2 SNPs on both chromosomes.

	s1	s2	s3	s4		sN
p1	0	1	0	0	0	1
p2	0	1	0	1	0	1
p3						
p4						
p5						
pM1	0	1	1	0	0	1

The H0 hypothesis is below.

	s1	s2	s3	s4		sN
p1	0	0	0	0	1	0
p2	0	0	0	1	0	0
p3	0	0	0	0	0	0
p4	0	0	0	0	0	0
p5	1	0	0	0	0	0
pM0	0	0	0	0	0	0
	$\sum_{j=1}^{M0} m_{1,j}$	$\sum_{j=1}^{M0} m_{2,j}$	$\sum_{j=1}^{M0} m_{3,j}$	$\sum_{j=1}^{M0} m_{4,j}$	$\sum_{j=1}^{M0} m_{i,j}$	$\sum_{j=1}^{M0} m_{N,j}$

Now there can be many ways to use this data. We briefly explain two.

Model 1: Probabilistic

In this case we measure for H0 and H1 the following:

$$P\left[s_{0,k} \mid H_0\right] = \frac{\sum_{k=1}^{M_0} m_{0,k}}{M_0}$$

and

$$P\left[s_{1,k} \mid H_1\right] = \frac{\sum_{k=1}^{M_1} m_{1,k}}{M_1}$$

These are the probabilities of the specific SNP given a disease state. We would like the probability of a disease state given the SNP. Namely:

$$P\left[H_1 \mid s_k\right] = \frac{P\left[s_k \mid H_1\right]}{P\left[s_1\right]} = \frac{P\left[s_k \mid H_1\right]}{P\left[s_1 \mid H_1\right]P\left[H_1\right] + P\left[s_1 \mid H_0\right]P\left[H_0\right]}$$

We can have estimates for all of these. Thus we can determine the probability from the data and choose the SNPs with the greatest probability. It must be pointed out that we should really include single or multiple SNPs, namely one chromosome or both. That will make the analysis a bit more complex.

Model 2: Clustering[12]

We can take the two data sets and create a hierarchical cluster. The metric for distance can any one of the many available. A possible algorithm would be to set all SNPs in H1 together with the largest number of positive SNPs and then do for example a nearest neighbor tree build up. We would do the same for H0 except the first cluster would have the minimal distance.

The tree result for clustering would go down to smaller or larger distances. When the two trees have the same distance we would trim off the remaining SNPs as just noise.

These are just two examples of examining SNPs.

However the SNP-SNP paper takes it another step. Then look at SNP pairs. Now we know that if we have N SNPs we have:

$${}_{N}C_2 = \frac{N!}{(N-2)!2!}$$ possible pairs.

For large N this becomes an explosive data set. We may look at the analysis in the same manner. We can look at every pair and calculate its probability and then select those which are largest. We can also do a cluster analysis to ascertain clusters which are the most prevalent in both classes.

[12] See Duda and Hart. Also see Cover and Hart for the original analysis.

But, and this is a critical but, the use of either will require judgment calls as when to taper or curtail an SNP pair set. As Lin et al state:

Our study demonstrated that frequent inconsistent results of individual SNPs may be partially due to SNP-SNP interactions. Similar SNP-SNP interaction patterns were observed in the majority of our results, but the individual SNP effects for the SNPs involved in the interactions could not be replicated in the two study groups.

The above statement is truly a statement of the basic fragility of this approach. Without any underlying physical connection one is relying on often unstable and arbitrary clustering algorithms for ascertaining prognostic measures which can have significant clinical effects, often presenting potential harm. They continue:

For the genetic association validation studies, it is well known the individual SNP results are difficult to reproduce. Hirschhorn et al. evaluated more than 600 reported associations and found less than 4% of the results were replicable among 166 associations that had been studied more than three times .

Again, the above statement is truly significant and a powerful reminder of the fact that just having data and putative correlations presents a challenge of prognostication. They continue:

Furthermore, the gene set identified in the main SNP effect and interaction approaches were totally different in our study. The four SNPs in the three genes (COL4A3, PDGFD and ELK3) with significant main effects in our two study groups did not overlap with the SNPs with significant interactions.

Thus, it is highly recommended to consider both main effects and interactions for comprehensively evaluating gene variations in genetic association studies. Our study findings , generated from the TRM approach by considering multiple SNPs simultaneously, may provide more useful information in building a multivariable prediction model than the pair-wise search approaches, which consider two SNPs at a time.

The results are still speculative. The authors present the following warning.

However, it should be noted that our study may not find all SNP-SNP interactions due to a limited sample size of each testing data set and characteristics of the TRM method. Although Random Forests have been shown to perform reasonably well in detecting pure SNP-SNP interactions, it still favors SNPs with strong main effects.

Therefore, although we have significant data, and correlative suspicions, we do not have a causative model.

4 OBSERVATIONS

Let us now review some of the findings and list out some of the key observations including the lingering questions.

4.1 SPECIFICS

As DeTours states, it may be all too easy to find aberrant genes, and even more so SNPs, independent of specific pathway models. And as we have argued, just within a pathway one may have a concern because it is also the intercellular signalling that is a concern as well. Even more so is the understanding of the process.

Specifically:

1. A melanocyte or prostate basal or prostate basal or luminal cell may be normal until something happens. What is it that happens, does a SNP occur in a somatic cell, why, when, and then what happens after that? Is the SNP a precursor, a subsequent event, or just of high coincidence?

2. If a SNP occurs, is that occurrence during the development of a DNA reading for protein generation or during cell replication. The opening of DNA for transcription may be the event which places the melanocyte at risk. If so then what is the risk process. Could it be radiation as suspected, or is it the next step in a Vogelstein like progression. Namely there may have already been SNP damages and this one could be the final straw. Is it a micro RNA problem? The dynamics of this are essential.

3. Knowing pathways, is it possible to work backward and determine what the aberrant change or changes were? Pathway changes are reflected by their products. What is some promoter protein is what changed, can that affect an SNP change as well?

4. What of the stem cell theory, must we look for the melanoma stem cell alone, and if so how can we identify it. The stem cell communicates, and that is a powerful mechanism to spread the cancer. How does it communicate and how is that related to the pathway.

However as Penney et al (2010) state:

No SNP reached genome-wide significance ($P \leq 1 \times 10^{-7}$); however, three independent SNPs had $P < 1 \times 10^{-5}$. One top-ranked SNP replicated ($P = 0.05$) in an independent follow-up study. Although no CNV had genome-wide significance, 14 CNVs showed nominal association with prostate cancer mortality ($P < 0.05$). No variants were significantly associated at a genome-wide level with prostate cancer mortality. Common genetic determinants of lethal prostate cancer are likely to have odds ratios <2.0. Genetic markers identified could provide biological insight to improve therapy for men with potentially fatal cancer. Larger studies are necessary to detect the genetic causes of prostate cancer mortality

The OR of less than 2.0 is frankly low. What should be a significant odds ratio? That is a more serious question.

Jin et al (2012) state:

Multiple prostate cancer (PCa) risk-related loci have been discovered by genome-wide association studies (GWAS) based on case–control designs. However, GWAS findings may be confounded by population stratification if cases and controls are inadvertently drawn from different genetic backgrounds.

In addition, since these loci were identified in cases with predominantly sporadic disease, little is known about their relationships with hereditary prostate cancer (HPC). The association between seventeen reported PCa susceptibility loci was evaluated with a family-based association test using 1,979 hereditary PCa families of European descent collected by members of the International Consortium for Prostate Cancer Genetics, with a total of 5,730 affected men.

The risk alleles for 8 of the 17 loci were significantly over-transmitted from parents to affected offspring, including SNPs residing in 8q24 (regions 1, 2 and 3), 10q11, 11q13, 17q12 (region 1), 17q24 and Xp11. In subgroup analyses, three loci, at 8q24 (regions 1 and 2) plus 17q12, were significantly over-transmitted in hereditary PCa families with five or more affected members, while loci at 3p12, 8q24 (region 2), 11q13, 17q12 (region 1), 17q24 and Xp11 were significantly over-transmitted in HPC families with an average age of diagnosis at 65 years or less.

Our results indicate that at least a subset of PCa risk-related loci identified by case–control GWAS are also associated with disease risk in HPC families.

Note that in Jin, we have added sets of SNPs, not the same. The Lin et al approach is to use CGEMS data set and then to utilize a data aggregating and clustering technique called MARS. Unfortunately, as with so many of these analyses, there is no way another reader could even come close to replicating the data from what has been presented. One must accept the answer as given, yet in my opinion highly suspect. Causality is missing and an underlying model as well. Albeit that Lin et al rely upon the angiogenesis argument, and that we have demonstrated that each of the purported genes may have merit, the problem is that the SNPs are in no way causal. There is no argument even remotely alleging that.

We now have a set of issues which should be addressed if these results have merit.

1. As regards to the SNPs; are they germline or somatic?

2. If somatic, what even gave to their rise?

3. If germline, what events turned them active if any?

4. What function does the SNP affect in the evolution of a malignant cell?

5. We understand the functions of the related genes, but what does the SNP related to the gene do to change the gene function, if any, and why?

4.2 MODELS AND WORLD VIEW

One of the issues we have argued at length is that one must be careful collecting correlative data in the absence of a world view model of what that data means in some interconnected manner. All too often we see massive amounts of data being correlated but not causative.

Let me return to the issue of the cancer stem cell, the CSC. This is a significant factor in many cancers and if it is as critical as is suspected then one must be careful as to what cells one is ascertain data on. As we have shown in our analysis of Prostate Cancer and HGPIN, one is told that HGPIN is a determinate precursor of PCa. However it has been observed that when high density prostate biopsies are performed, 24 or more cores, and HGPIN is observed, that upon a second and subsequent biopsies the HGPIN has not become a PCa but in fact has regressed totally. One explanation is the removal of the CSC from the prostate. As Navin and Hicks state:

In the late 1990's an alternative model emerged that challenged the primary assumption of the previous models by assuming that only a minority of tumor cells could proliferate indefinitely. The cancer stem cell (CSC) hypothesis became widely accepted as the leading model for tumor progression. The CSC hypothesis posits that a rare population of stem cells within the solid tumor is the only subpopulation with the ability for unlimited proliferation (Fig. 1e). The model assumes: (1) a rare population of cancer stem cells proliferate indefinitely, (2) the majority of tumor cells have limited proliferation, and (3) the rare cells continuously give rise to the major population. Cancer stem cells were originally believed to arise from normal stem cells, but it is now thought that any somatic cell may become a cancer stem cell.

Now the authors examine the many models of cancer. These are systemic models wherein one can validate them by taking measurements. They continue:

Biological models are by definition built upon incomplete information. At best, these explicit models for tumor progression provide guideposts for further exploration. As technology continues to evolve, the analysis of cancer samples of complex mixtures will give way to methods aimed at the individual cell. Such methods will enable single cancer cells to be tracked as they progress to form the primary tumor and traced as they migrate through the body to seed the metastasis. In the near future the cost of deep sequencing a mammalian genome, whether from a tumor sample or a few disseminated cells will be approximately equivalent to the current price of a microarray experiment.

Single cell genomes are also ideal for constructing detailed lineages of tumor progression, because individual mutations in a genome can be traced as they are inherited and expanded in subpopulations. As we bring the magnifying glass closer, we may also be able to track the genetic stepping stones for tumor growth, or follow the genetic changes in circulating tumor cells as they progress from the primary to metastasis. Perhaps, we will find evidence that individual circulating tumor cells return to the primary tumor after developing offsite as the self-

seeding model suggests. It is then that these predictive genetic models will have realized their full value.

Now the following two quotes are essential in terms of understanding what come first; the model or the data.

"It is also a good rule not to put too much confidence in the observational results that are put forward until they are confirmed by theory." Arthur Eddington

"It is quite wrong to try founding a theory on observable magnitudes alone. It is the theory which decides what we can observe." Albert Einstein

Thus we argue that studies such as the ones discussed herein are devoid of models or world views and may be nothing more than happenstance. That is a concern.

5 REFERENCES

5.1 WEB SITES

http://www.cancer.gov/cancertopics/understandingcancer/geneticvariation/AllPages

http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml

http://www.biotechniques.com/news/341609SNPping-Away-at-Cancer/biotechniques-341608.html?autnID=323573

http://www.ncbi.nlm.nih.gov/books/NBK21088/ (SNP Data Base)

http://www.ncbi.nlm.nih.gov/projects/SNP/batchquery.html SNPdb

http://www.ncbi.nlm.nih.gov/projects/SNP/docs/dbSNP_VCF_Submission.pdf VCF Form

http://www.ncbi.nlm.nih.gov/books/NBK44420/ Finding SNPs

http://www.ncbi.nlm.nih.gov/books/NBK3848/ SNP Help Manual

5.2 DOCUMENTS

1. Azzato, E., et al, Common germ-line polymorphism of C1QA and breast cancer survival, British Journal of Cancer (2010) 102, 1294 – 1299.
2. Barrett, J., et al, Genome-wide association study identifies three new melanoma susceptibility loci, Nature Genetics, 43, pp 1108–1113, (2011), http://www.nature.com/ng/journal/v43/n11/pdf/ng.959.pdf
3. Bauer, A., C. Stratakis, The lentiginoses: cutaneous markers of systemic disease and a window to new aspects of tumorigenesis, www.jmedgenet.com , J Med Genet 2005;42:801–810.
4. Cifola, I, et al, Comprehensive Genomic Characterization of Cutaneous Malignant Melanoma Cell Lines Derived from Metastatic Lesions by Whole-Exome Sequencing and SNP Array Profiling, PLOS ONE, 1 May 2013,Volume 8, Issue 5. http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0063597&representation=PDF
5. Cover, T., P. Hart, Nearest Neighbor Pattern Classification, IEEE Info Theo, January 1967.
6. Datta, M., et al, Perlecan, a candidate gene for the CAPB locus, regulates prostate cancer cell growth via the Sonic Hedgehog pathway, Molecular Cancer2006, 5:9 doi:10.1186/1476-4598-5-9.

7. Delassus, G., et al, New pathway links from cancer-progression determinants to gene expression of matrix metalloproteinases in breast cancer cells, J Cell Physiol. 2008 Dec; 217(3):739-44. doi: 10.1002/jcp.21548.
8. Detours, V., Confounded Cancer Markers, http://www.the-scientist.com/?articles.view/articleNo/31482/title/Opinion--Confounded-Cancer-Markers/ The Scientist, December 7, 2011.
9. Dey, A., et al, Colony-stimulating Factor-1 Receptor Utilizes Multiple Signaling Pathways to Induce Cyclin D2 Expression, Mol Biol Cell. 2000 November; 11(11): 3835–3848.
10. Dickinson, I., W. Duncan, The SLIT/ROBO pathway: a regulator of cell function with implications for the reproductive system, Reproduction. 2010 April ; 139(4): 697–704.
11. Duda, R., P. Hart, Pattern Classification and Scene Analysis, Wiley (New York) 1973.
12. Dunson, D., et al, Bayesian Selection and Clustering of Polymorphisms in Functionally-Related Genes, to be published.
13. Eeles, R., et al,, Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array, Nature Genetics, April, 2013.
14. Evans, C., AUA 2010 - Are newly described prostate cancer susceptibility variants associated with prostate cancer aggressiveness? - Session Highlights http://www.urotoday.com 25 August, 2010.
15. Foulkes, W., Inherited Susceptibility to Common Cancers, NEJM, 359;20 www.nejm.org November 13, 2008
16. Grochoula, L. et al, Single-nucleotide Polymorphisms in the p53 Signaling Pathway, Cold Spring Harb Perspect Biol 2010.
17. Helfand, B., et al, Personalized Prostate Specific Antigen Testing Using Genetic Variants May Reduce Unnecessary Prostate Biopsies, http://www.jurology.com/article/S0022-5347%2812%2905850-8/abstract , The Journal of Urology, Volume 189, Issue 5 , Pages 1697-1701, May 2013
18. Helfand BT, Kan D, Modi P, et al. Prostate cancer risk alleles significantly improve disease detection and are associated with aggressive features in patients with a “normal” prostate specific antigen and digital rectal examination. Prostate. 2011;71(4):394-402. PMID:20860009
19. Hodis, E., et al, A Landscape of Driver Mutations in Melanoma, Cell 150, 251–263, July 20, 2012. http://211.144.68.84:9998/91keshi/Public/File/42/150-2/pdf/1-s2.0-S0092867412007787-main.pdf
20. Ickstadt, K., et al, Cluster Analysis: A Comparison of Different Similarity Measures for SNP Data, Econstor, 2005.
21. Jia, P., et al, Integrative pathway analysis of genome-wide association studies and gene expression data in prostate cancer, BMC Systems Biology 2012, 6(Suppl 3):S13
22. Jin G., et al, Validation of prostate cancer risk-related loci identified from genome-wide association studies using family-based association analysis: evidence from the International

Consortium for Prostate Cancer Genetics (ICPCG), Hum Genet. 2012 July ; 131(7): 1095–1103.

23. Jin, G., et al, Genome-wide Association Study Identifies Loci at ATF7IP and KLK2 Associated with Percentage of Circulating Free PSA, Neoplasia, Volume 15 Number 1 January 2013 pp. 95–101
24. Jin, G., et al, Validation of prostate cancer risk-related loci identified from genome-wide association studies using family-based association analysis: evidence from the International Consortium for Prostate Cancer Genetics (ICPCG), Hum Genet. 2012 July ; 131(7): 1095–1103.
25. Lin, D., et al, Genetic Variants in the LEPR, CRY1, RNASEL, IL4, and ARVCF Genes Are Prognostic Markers of Prostate Cancer-Specific Mortality, Published OnlineFirst August 16, 2011; doi: 10.1158/1055-9965.EPI-11-0236. http://cebp.aacrjournals.org/content/early/2011/08/12/1055-9965.EPI-11-0236.abstract
26. Lin, H. SNP-SNP Interaction Network in Angiogenesis Genes Associated with Prostate Cancer Aggressiveness, PLOS ONE, www.plosone.org , 1 April 2013, Volume 8, Issue 4, e59688.
27. Liu, H., et al, Association between functional polymorphisms in genes involved in the MAPK signaling pathways and cutaneous melanoma risk, Oxford Journals, Life Sciences & Medicine, Carcinogenesis, Volume 34, Issue 4, Pp. 885-892, http://carcin.oxfordjournals.org/content/34/4/885.abstract
28. Maeshima, Y., Angiogenesis and Cancer, Springer, www.springer.com/cda/content/.../cda.../9781588296771-c2.pdf
29. McGarty, T.P., Melanoma Genomics, Draft 2013, http://www.telmarc.com/Documents/Books/Melanoma%20Genomics%2007.pdf
30. McGarty, T. P., Prostate Cancer Genomics, Draft 2013, http://www.telmarc.com/Documents/Books/Prostate%20Cancer%20Systems%20Approach%2003.pdf
31. McGarty, Extracellular Matrix vs. Intracellular Pathways , Telmarc White Paper No 88, http://www.telmarc.com/Documents/White%20Papers/88%20Extracellular%20Matrix%2002.pdf
32. Muller, T., et al, Cluster Analysis: A Comparison of Different Similarity Measures for SNP Data, Econstar, 2005.
33. Murphy, K., Machine Learning, MIT Press (Cambridge) 2012.
34. Nagase, H., et al, Structure and function of matrix metalloproteinases and TIMPs, Cardiovascular Research 69 (2006) 562 – 573.
35. Navin, N., J. Hicks, Tracing the Tumor Lineage, Mol Oncol. 2010 June ; 4(3): 267–283
36. Pashayan N., et al, Polygenic susceptibility to prostate and breast cancer: implications for personalised screening, British Journal of Cancer (2011) 104, 1656–1663.
37. Penney, K., et al, Association of KLK3 (PSA) genetic variants with prostate cancer risk and PSA levels, Carcinogenesis vol.32 no.6 pp.853–859, 2011.

38. Penney, K., Genome-wide Association Study of Prostate Cancer Mortality, Cancer Epidemiol Biomarkers Prev 2010;19:2869-2876.
39. Schulterman, M., A Novel Function for Fibulin-5 in Controlling Integrin Induced Production: Implications on Angiogenesis and Cancer, PhD Thesis, The University of Texas Southwestern Medical Center at Dallas, 2009.
40. Selinski, S., K. Ickstadt, Similarity Measures for Clustering SNP Data, to be published.
41. Speicher, M., et al, Human Genetics, Springer (New York) 2010.
42. Syvanen, A., Accessing Genetic Variation: Genotyping Single Nucleotide Polymorphisms, Nature, 930, DECEMBER 2001, VOLUME 2.
43. Wang, L. et al, Genomics and Drug Response, NEJM, 364;12 nejm.1144 org march 24, 2011.
44. Zheng, S., et al, Cumulative Association of Five Genetic Variants with Prostate Cancer, NEJM, February 28, 2008.
45. Wakeley, J., et al, The Discovery of Single-Nucleotide Polymorphisms—and Inferences about Human Demographic History, Am. J. Hum. Genet. 69:1332–1347, 2001.

6 ABBREVIATIONS:

1. AMFS Australian Melanoma Family Study
2. CI confidence interval
3. CM cutaneous melanoma
4. DUSP Dual-specificity phosphatase
5. ERK extracellular signal-regulated kinase
6. FDR false discovery rate
7. GWAS genome-wide association study
8. JNK c-jun N-terminal kinase
9. LD linkage disequilibrium
10. MAF minor allele frequency
11. MAPK mitogen-activated protein kinase
12. miRNA micro RNA
13. mRNA messenger RNA
14. OR odds ratio
15. QC quality control
16. Q-MEGA Queensland Study of Melanoma, Environment and Genetic Associations
17. SNP single nucleotide polymorphism
18. VEGAS Versatile Gene-Based Test for Genome-wide Association.