CNVs and Prostate Cancer

Terrence P McGarty
White Paper No 120
December, 2014

There have been a few recent papers proposing the use of Copy Number Variations (CNVs) as prognostic for Prostate Cancer, PCa. We examine these papers herein and present several issues which should warrant further examination. We also examine the Press response to such papers and pose questions on the duty of the researchers in mollifying the Press exuberance in its descriptions.

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.

Contents

1 INTRODUCTION

There is a continuing flow on putative markers for both diagnostic and prognostic studies on human tissues, especially for prostate cancer (PCa). In a recent study there was a focus on Copy Number Variations ("CNV"s) that were touted as significant prognostic markers. We examine the CNV markers and we use this recent study as another example of a dissonance and disconnect between what the Press states and what apparently is the true case. Mikropoulos et al present a recent summary of many of the genomics based tests which apply. However the focus here is CNV tests.

We will specifically focus on a recent Lancet paper by Lelonde et al which presents an analysis of several hundred patients who had prostate cancer and examined them over a period of time for survival. The analysis focused on CNVs which became present in the PCa cells. The authors then proposed that the presence of CNVs across certain genome spaces were highly prognostic of good or poor outcomes. Furthermore the Press seems to take the results and make statements which in my opinion far outstretch the results presented. Our goal herein is twofold. First the paper presents an interesting platform to examine the current state of using CNVs as a means for PCa prognostics. Secondly, and equally important, is an analysis of how the medical press all too often makes statements which seem to exceed the results contained in the published results.

CNV are variations in a gene that are deletions, insertions, inversions, copy number variants or sequential duplications[1]. Now these CNVs are present in over 12% of the genome or some 360 million nucleotides. Most of this is benign, some is problematic. If we look at malignant cells as compared to benign somatic cells we may also observe CNVs in the malignant cells that were not in the benign cells. That in essence is what we will discuss herein. Specifically if we examine prostate cancer (PCa) and we look for CNV variations between putatively benign cells of a patient and the clearly malignant cells, we see certain CNV patters on certain genes. It is then argued that this is pathognomonic for some specific prognostic state. The technique used to compare these two gene states, the benign somatic cell and the PCa cell, is called array comparative genome hybridization (aCGH). We shall examine this technique as well. It is simply a technique that uses the tagging of cells with florescent markers and then does so for benign and malignant cells, marking the DNA, and then comparing them with some analytical methods. The result is the presentation of CNV regions.

However, this observation is not clearly causal. It may be correlative however recent papers have effectively argued it has substantial prognostic value. Let us begin by a Medscape article discussing the paper in question[2]:

[1] See Redon et al.

[2] http://www.medscape.com/viewarticle/835051_print

A novel genetic test based on a prostate cancer's unique genetic fingerprint can accurately predict disease recurrence in men undergoing local treatment and distinguish those who will require additional systemic therapy from those who will not, investigators report

Now this opening statement posits that the test can accurately demonstrate who will and will not benefit from treatment. Frankly this is far from the truth when one reads the paper in detail. A few hundred patients in a small trial and an AUC (area under the curve of the ROC or receiver operating characteristics) of 0.7 at best is hardly definitive. The article continues:

This study adds to a "burgeoning explosion" of so-called biomarkers to help clinicians who care for patients with prostate cancer to more precisely assess risk and direct treatment programs, said Marc Garnick, MD, professor of medicine at the Beth Israel Deaconess Medical Center and Harvard Medical School in Boston, and editor-in-chief of the Harvard Medical School Annual Report on Prostate Diseases. "The current measurements evaluate both unstable genes and the presence of measurements of prostate cancer hypoxia, both of which, if present, can enable a more aggressive behavior of prostate cancer," he told Medscape Medical News. ...

Dr Bristow and colleagues exhaustively report on complicated measurements — not widely available — and then retrospectively assess the clinical behavior of the patients based on the presence or absence of these biomarkers, Dr Garnick observed. "The results seem to identify a subset of patients whose cancers may behave less or more aggressively," he added. "That is the exciting aspect."

However, Dr Garnick cautioned that as a clinician caring for patients with prostate cancer, some of the findings are confusing, especially the disconnect between the new biomarkers and the traditional bedside parameters of prostate-specific antigen, Gleason score, and T stage — the key ways in which clinical decision-making is currently generated. "While it is entirely too premature to even consider using these measurements in clinical practice, in the future, these, along with a multitude of other tests, may help by adding precision to the selection of treatments for high-risk patients," Dr Garnick said. Current treatments will also have to improve if these emerging biomarker tests are to ever reach their full potential, he added.

Then in a Healio report they state[3]:

Researchers in Canada have developed a genomic test to identify patients with prostate cancer who are at high risk for recurrence after surgery or radiotherapy, according to findings presented at the European Society for Radiotherapy and Oncology.

Again we argue that the definitive statement is not correct given the data reported in the paper in Lancet. In a Medical Daily release, done in April well before publication, they state[4]:

[3] http://www.healio.com/hematology-oncology/prostate-cancer/news/onlin

[4] http://www.medicaldaily.com/will-my-prostate-cancer-return-new-genetic-test-detects-men-likely-have-recurring-cancer-274566

Over the next two to three years, the researchers will work to validate the test in different and larger groups of patients. "If all goes well, then this will lead to a new test for cancer patients that can be turned around in three days and will tell doctors which patients will do well with local treatment alone — surgery or radiotherapy — and which will need extra treatment."

Note the tentative nature of the statement. It is not clear that between ASCO in April 2014 and the Lancet article in November 2014 that there were additional definitive results.

The questions discussed herein are as follows:

1. What value do CNVs provide for PCa prognostic value? Namely, regardless of the results, which are of substantial interest, can CNV testing be truly prognostic? The issue is that the researchers looked at changes in CNV between putatively benign cells from the patient and those in the prostate cancer. Thus do we learn anything from this technique that helps?

2. Given a prognostic data result, even with a 0.7 AUC, do we have treatment for patients with poorer prognostic values? Generally the answer is no. It is no because we have no clear definitive nexus between the tests result, the cause of the cancer, and a therapeutic that incapacitates the cause. Thus we have no target for say a kinase inhibitor. Thus we ask, what is the value of the test?

3. What responsibility do researchers have in dealing with the Press? The Press seeks items that draw attention since in drawing attention they generate advertising revenue. Yet at the same time patients, even more so than physicians, see the results and reach less than accurate conclusions. They may engender hope where none is. Thus ethically is this appropriate?

4. There are two elements of the study. One was the selection of patients with PCa and the determination of their survivability based on specific CNV profiles. The second is the use of aCGH technique to make this comparison. The first element is adding just another metric into an already well occupied set of tools. The second, however, is an attractive tool for examining cell changes. aCGH is worth study on its own.

2 THE STUDY

We begin with a brief review of the Lancet study in question. We will get into the specifics in a short while but it is worth the overview to see what is being claimed and what the bases for the claims are. From the article itself we have:

A signature to classify patients as potential responders or non-responders to local therapy would have great clinical use if it was treatment-independent (ie, effective both for patients undergoing radical prostatectomy or image-guided radiotherapy) and could be done on initial diagnostic biopsies. Such a signature could triage patients at greatest risk of failure into clinical trials for treatment intensification and justify potential added toxic effects.

Indeed a classifier is always of use. Yet as we indicated above the use implies that with the knowledge we may be able to treat the separated groups. In this case however one group will do just fine and the other will die. This is independent of any available treatment. In addition with an AUC of only 0.7 we still have substantial uncertainty.

DNA copy number alterations in PTEN, NKX3-1, MYC, and STAR are associated with adverse prognosis, and RNA-based gene signatures might differentiate indolent and non-indolent, lethal prostate cancer. TMPRSS2–ERG fusion status does not predict prognosis after radical prostatectomy or image-guided radiotherapy. Importantly, tumour cells exist within a heterogeneous tumour microenvironment with dynamic gradients of hypoxia that have been linked to metastatic potential. Indeed, patients with prostate cancer with hypoxic tumours rapidly fail treatment (eg, within 2 years) after radical prostatectomy or image-guided radiotherapy.

The CNV coverage of key genes is important. Indeed it is worth reading the more than 120 pages of added detail in the Lancet article to see that much of the results have such value.

Biochemical relapse was associated with indices of tumour hypoxia, genomic instability, and genomic subtypes based on multivariate analyses. We identified four genomic subtypes for prostate cancer, which had different 5-year biochemical relapse-free survival.

Genomic instability is prognostic for relapse … Up to now, the interplay of genomic instability and tumour microenvironment in modulation of treatment outcome has been unexplored. We therefore aimed to develop clinically relevant prognostic indices, with use of integrated tumour DNA and microenvironmental indices, to robustly predict patient outcome.

The results seem to imply that if one looks at the genes mentioned, key genes known to be associated with PCa that any CNV disturbance results in a deteriorated prognostic result.

We list some details on the four above mentioned genes below. Not that three are on chromosome 8 and PTEN is on chromosome 10. PTEN and MYC are well known to have relationships to cancer as does NKX3-1 in PCa. Thus it would seem logical that any CNV that changes the expression of these genes is problematic.

Gene	***Location***	***Function***
PTEN	10q23.3	*…tumor suppressor that is mutated in a large number of cancers at high frequency*
NKX3-1	8p21.2	*…homeobox-containing transcription factor. This transcription factor functions as a negative regulator of epithelial cell growth in prostate tissue*
MYC	8q24.21	*…multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation. It functions as a transcription factor that regulates transcription of specific target genes*
STAR	8p11.2	*….acute regulation of steroid hormone synthesis by enhancing the conversion of cholesterol into pregnenolone*

The analysis in the Lancet paper performed the following:

DNA was extracted from pretreatment biopsies that consisted of at least 70% tumour cells as estimated by a pathologist, and a custom array was used to detect copy number alterations. Intraglandular measurements of partial oxygen pressure were taken before radiotherapy with an ultrasound-guided transrectal needle piezoelectrode…

We developed four prognostic indices and validated them for prediction of biochemical relapse (appendix). The appendix provides an overview of our approach to develop treatment-independent, integrated prognostic indices.

First, we identified unique genomic subtypes with use of unsupervised hierarchical clustering.

Second, we used the percentage of a patient's genome harbouring copy number alterations (percent genome alteration) as a surrogate for genomic instability, and assessed this proportion together with tumour hypoxia.

Third, we undertook supervised machine learning with a random forest to develop a statistical model, resulting in a DNA signature, which classified patients at risk of biochemical relapse on the basis of their copy-number profiles.

We compared the resulting signature with published RNA based signatures.

Now it is useful to examine similar CNV and PCa results specifically those by Hieronymus et al in PNAS in July 2014. They state:

We find that CNA burden across the genome, defined as the percentage of the tumor genome affected by CNA, was associated with biochemical recurrence and metastasis after surgery in these two cohorts, independent of the prostate-specific antigen biomarker or Gleason grade, a major existing histopathological prognostic variable in prostate cancer. Moreover, CNA burden was associated with biochemical recurrence in intermediate-risk Gleason 7 prostate cancers, independent of prostate-specific antigen or nomogram score. We further demonstrate that CNA

burden can be measured in diagnostic needle biopsies using low-input wholegenome sequencing, setting the stage for studies of prognostic impact in conservatively treated cohorts. …

They then conclude:

The discovery of molecular biomarkers for prostate cancer has been hindered by the paucity of molecular subtypes with distinct outcomes. Part of the difficulty stems from the long natural history of prostate cancer and its low rate of progression. As a result, whereas many comprehensive genomic studies have defined alterations in prostate cancer genomes, none have had sufficient clinical and outcome annotation to generate prognostic advances. We have collected the clinically annotated cohorts presented here to help address these challenges. The initial cohort has been used extensively by others for prognostic discovery and validation (7, 8, 27) and is updated here by providing a definitive clinical endpoint, metastasis, in significant numbers.

We have shown that CNA burden, as a global measure of the level of CNA across tumor genomes, is associated with BCR over a broad spectrum of clinical presentations and adds additional information to currently available clinicopathological variables. CNA burden is also associated with the definitive endpoint of metastasis. Moreover, CNA burden is significantly associated with BCR in intermediate risk Gleason 7 patients[5]*, raising the possibility of better stratifying this intermediate risk population. CNA burden varies significantly in other cancers, with many showing large ranges in CNA burden across their populations. It may therefore be fruitful to also explore the prognostic significance of CNA burden in other cancers as well.*

Thus the Lancet paper does somewhat reiterate the NAS paper and also the results broadly depict a silencing or expression alteration to a set of well-known genes. As we shall note, one of the key questions is; why and how does this happen?

[5] BCR is defined as an increase of PSA of >0.2 ng/mL on two occasions.

3 CNV

Each day we see more relationships between genes, SNPs, miRNA, lncRNA, STR and now CNVs to some form of cancer. There is a recent paper in The American Journal of Pathology (2012) which relates CNVs to prostate cancer, PCa, and the prognosis of the disease. In that paper we find many prior works efforts as well. Thus examining CNVs and looking at them as prognostic is of some import.

3.1 CNV DESCRIPTIONS

We start with a brief discussion of a CNV. It is defined as follows:

Copy number variant (CNV): A duplication or deletion event involving >1 kb of DNA.

Simply a CNV may be the addition or deletion or movement of one or more copies of a gene or part thereof in a chromosome. It simply adds to the chromosome. They are quite common and thus are seen frequently. Some are related to certain genetically inherited disorders. In the paper at point they are used to ascertain potentially prognostic data.

From the paper by Yu et al[6]:

The prediction of prostate cancer clinical outcome remains a major challenge after the diagnosis, even with improved early detection by prostate-specific antigen (PSA) monitoring.

To evaluate whether copy number variation (CNV) of the genomes in prostate cancer tumor, in benign prostate tissues adjacent to the tumor (AT), and in the blood of patients with prostate cancer predicts biochemical (PSA) relapse and the kinetics of relapse, 241 samples (104 tumor, 49 matched AT, 85 matched blood, and 3 cell lines) were analyzed using Affymetrix SNP 6.0 chips.

By using gene-specific CNV from tumor, the genome model correctly predicted 73% (receiver operating characteristic P = 0.003) cases for relapse and 75% (P < 0.001) cases for short PSA doubling time (PSADT, <4 months). The gene-specific CNV model from AT correctly predicted 67% (P = 0.041) cases for relapse and 77% (P = 0.015) cases for short PSADT. By using median-sized CNV from blood, the genome model correctly predicted 81% (P < 0.001) cases for relapse and 69% (P = 0.001) cases for short PSADT.

By using median-sized CNV from tumor, the genome model correctly predicted 75% (P < 0.001) cases for relapse and 80% (P < 0.001) cases for short PSADT. For the first time, our analysis

[6] Yu, Y., et al, Genome Abnormalities Precede Prostate Cancer and Predict Clinical Relapse, The American Journal of Pathology - June 2012 (Vol. 180, Issue 6, Pages 2240-2248, DOI: 10.1016/j.ajpath.2012.03.008). http://www.journals.elsevierhealth.com/periodicals/ajpa/article/S0002-9440%2812%2900241-6/abstract

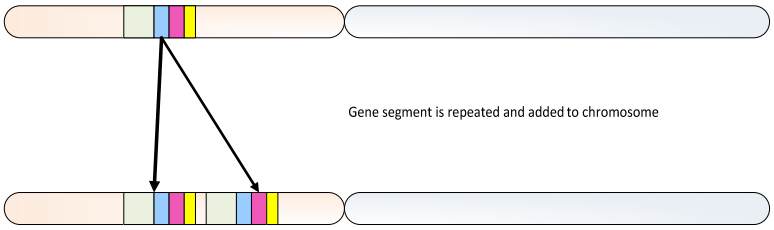
indicates that genomic abnormalities in either benign or malignant tissues are predictive of the clinical outcome of a malignancy.

We briefly examine the CNV in general. In the work of Freeman et al we have[7]:

DNA copy number variation has long been associated with specific chromosomal rearrangements and genomic disorders, but its ubiquity in mammalian genomes was not fully realized until recently. Although our understanding of the extent of this variation is still developing, it seems likely that, at least in humans, copy number variants (CNVs) account for a substantial amount of genetic variation.

Since many CNVs include genes that result in differential levels of gene expression, CNVs may account for a significant proportion of normal phenotypic variation. Current efforts are directed toward a more comprehensive cataloging and characterization of CNVs that will provide the basis for determining how genomic diversity impacts biological function, evolution, and common human diseases.

We show an example of a CNV below graphically.



Here we have depicted a gene, the multicolor object in a chromosome and we have shown a CNV with an identical copy of the gene in the same chromosome. The authors continue:

CNVs often occur in regions reported to contain, or be flanked by, large homologous repeats or segmental duplications. Segmental duplications could arise by tandem repetition of a DNA segment followed by subsequent rearrangements that place the duplicated copies at different chromosomal loci. Alternatively, segmental duplications could arise via a duplicative transposition-like process: copying a genomic fragment while transposing it from one location to another

It must be noted that these are identical duplications of the genes, or segments thereof. If of a gene the segment can be transcribed as easily as the original. This raises the question that the resulting translated protein is at a potential multiple level of concentration, although this may not necessarily ne the case. They continue:

Large duplications and deletions have been known for some time to be related to the presentation of specific genetic disorders, presumably as a result of copy number changes

[7] Freeman, J., Copy number variation: New insights in genome diversity, Published in Advance June 29, 2006, doi: 10.1101/gr.3677206 *Genome Res. 2006. 16: 949-961* http://genome.cshlp.org/content/16/8/949.full.html#ref-list-1

involving dosage-sensitive developmental genes. This has led to the establishment of genetic diagnostic tests for certain, well-characterized microdeletion and microduplication syndromes (e.g., Angelman syndrome, DiGeorge syndrome, Charcot-Marie-Tooth disease, etc.).

If a de novo chromosomal aberration is recognized in a patient with a constitutional genetic abnormality (i.e., follow-up studies fail to reveal a similar chromosomal aberration in either of the two parents, and non-paternity has been excluded) and the aberration is not one of the dozen or so well-known common chromosomal polymorphisms (e.g., inversion on chromosome 9), the aberration is assumed to be the cause of the clinically recognized abnormal phenotype.

Finally the CNVs are not necessarily related to disorders. Some have CNV but many CNV are not noticeable. They thus state:

CNVs that do not directly result in early onset, highly penetrant genomic disorders may consequently be considered to be neutral in function, but afterward shown to play a role in later onset genomic disorders or common diseases. Analyses of the functional attributes of currently known CNVs reveal a remarkable enrichment for genes that are relevant to molecular–environmental interactions and influence our response to specific environmental stimuli.

These include, but are not limited to, processes involving drug detoxification (e.g., glutathione-S-transferase, cytochrome P450 genes, and carboxylesterase gene families), immune response and inflammation (e.g., leukocyte immunoglobulin-like receptor, defensin, and APOBEC gene families), surface integrity (e.g., late epidermal cornified envelope and mucin gene families), and surface antigens (e.g., galectin, melanoma antigen gene, and rhesus blood group gene families). Likewise, some CNVs encompass genes that may contribute to interindividual variation in drug responses, as well as in immune defense and disease resistance/susceptibility among humans.

From the Thorne and District Gazette[8]:

This study was appropriately designed to see whether patients who have different outcomes have differences in copy number variation. However, before this technique can be used as a test, it will have to be trialled on a much larger cohort of people, so that researchers can get a clearer picture of its use in clinical settings. For example, researchers will need to know how often the test might miss patients that are likely to relapse, and also how often the test incorrectly suggests a person's cancer is likely to relapse, which could lead them to have unnecessary further treatment. Also, as the authors note, the techniques used in this study need high-quality DNA, so may be difficult and expensive to perform…

The article then states regarding the outcomes:

1. *Approximately one-third of the patients had a relapse soon after surgery, with a median time to progression of 1.9 months.*
2. *One-third had a relapse but much more slowly, with a median time to progression of 47.4 months.*
3. *One-third of patients in the cohort were free of cancer for at least five years.*

[8] http://www.thornegazette.co.uk/news/health/behind-the-headlines/dna-blood-test-for-prostate-cancer-1-4606100

Based on the associations they found, the researchers developed an algorithm for predicting whether a patient would relapse, and how quickly they would relapse. This was based on whether the genetic code at specific locations was repeated or deleted, or on the size of copy number variation found across a person's genome. They then tested their prediction model on an additional 25 samples.

They then conclude:

The researchers found that the prostate cancer samples had a large number of genetic abnormalities.

(i) Deletions of specific regions occurred at high frequency, and amplification (abnormal repetitions) of other regions occurred in a subset of samples.

(ii) Healthy tissue adjacent to a tumour also had similar amplification and deletion patterns.

(iii) The blood of patients with prostate cancer also contained copy number variations, and some of these variations occurred in the same locations within the DNA as they had in the prostate cancer samples.

The researchers then developed a tool to predict whether a cancer would relapse based on DNA regions that had a significant proportion of amplification or deletion in prostate tissue samples from patients who relapsed, but not in patients who did not relapse. The prediction model looking at cancer tissue samples could predict a relapse correctly 73% of the time.

(i) It had a 75% accuracy for predicting rapid relapse.

(ii) The prediction model based on examining healthy tissue samples could predict a relapse 67% of the time.

(iii) It had a 77% accuracy for predicting rapid relapse.

(iv) This blood-based prediction model had an accuracy of 81% for predicting relapse, and a 69% accuracy for predicting rapid relapse.

(v) The cancer tissue analysis tool had an accuracy of 70% for predicting relapse, and 80% for rapid relapse.

(vi) The healthy tissue sample tool had an accuracy of 70% for relapse and rapid relapse, and

(vii) the blood sample tool had an accuracy of 100% for relapse and 80% for rapid relapse.

This is but another way to examine PCa cells. It does pose several questions:

1. Pathways: Is there also a set of pathway malfunctions that one sees in PCa also present here?

2. Is the CNV an artifact or causative. If causative then what is the specific process and how does it relate to known pathways.

3. This is a complex cellular measurement of genes. Is this cost effective?

4. The classic issue of stem cells again is raised. What chromosomes do we look at? Is this specific only to the PCa cells, the PCa stem cells, and all cells?

Now we can also compare CNVs and SNPs. We use a modified Table from Mikhail[9]:

	Genomic rearrangements (including CNVs)	***Base pair (bp) alterations***
Size	• Thousands to millions of bp	• Small scale gene mutations (e.g. point mutations)
Gene content	• One to several genes	• One gene
Molecular mechanism	• Mechanisms mediated or stimulated by genomic architecture • OR • Exogenous factors (e.g. ionizing radiation)	• Errors of DNA replication and/or repair • OR • Exogenous factors (e.g. chemical mutagens)
Locus-specific mutation rate (μ)	• CNVs: 1.7x1o-6 - 1.2x1Q-4	• Single-nucleotide changes: 1.8 - 2.5x1 o-8
Method of detection	• G-banded chromosomes • FISH • Cytogenomic arrays	• DNA sequencing • Other molecular techniques

3.2 CNV ELEMENTS

We now want to examine some of the details regarding the elements of CNV applications to PCa. From Redon:

Deletions, insertions, duplications and complex multi-site variants, collectively termed copy number variations (CNVs) or copy number polymorphisms (CNPs), are found in all humans10 and other mammals examined. We defined a CNV as a DNA segment that is 1 kb or larger and present at variable copy number in comparison with a reference genome10. A CNV can be simple in structure, such as tandem duplication, or may involve complex gains or losses of homologous sequences at multiple sites in the genome …

[9] See Mikhail, Genetics and Genomics Clinical Research Center, October 2013. http://www.uab.edu/hcgs/images/PDF_documents/Fall2013ImmersionCourse/Copy_Number_Variation.pdf

A total of 1,447 copy number variable regions (CNVRs), which can encompass overlapping or adjacent gains or losses, covering 360 megabases (12% of the genome) were identified in these populations. These CNVRs contained hundreds of genes, disease loci, functional elements and segmental duplications. Notably, the CNVRs encompassed more nucleotide content per genome than SNPs, underscoring the importance of CNV in genetic diversity and evolution. The data obtained delineate linkage disequilibrium patterns for many CNVs, and reveal marked variation in copy number among populations.

From Miller et al we have the following Table (modified) which details the use of CNVs in many applications:

Category	*General Findings*	*Pathogenic*	*Benign*
1	a. Identical CNV inherited from a healthy parent		X
	b. Expanded or altered CNV inherited from a parent	X	
	c. Identical CNV inherited from an affected parent	X	
2	a. Similar to a CNV in healthy relative		X
	b. Similar to a CNV in an affected relative	X	
3	CNV is completely contained within genomic imbalance defined by a high-resolution technology in a CNV database of healthy individuals		X
4	4. CNV overlaps a genomic imbalance defined by a high-resolution technology in a CNV database for patients with ID/DD, ASD, or MCA	X	
5	CNV overlaps genomic coordinates for a known genomic-imbalance syndrome (i.e., previously published or well-recognized deletion or duplication syndrome)	X	
6	CNV contains morbid OMIM genes	X	
7	a. CNV is gene rich	X	
	b. CNV is gene poor		X
	General Findings		
1	a. CNV is a deletion	X	
	b. CNV is a homozygous deletion	X	
2	a. CNV is a duplication (no known dosage sensitive genes)		X
	b. CNV is an amplification (greater than 1 copy gain)	X	
3	CNV is devoid of known regulatory elements		X

3.3 SNPS AND OTHER MARKERS

CNVs have been used as markers in a broad base of areas. SNPs also have been used. Whereas CNVs are additions or deletions of 1KB or more, SNPs are single base changes. An SNP may be in an intron or exon, non-coding or coding region, and a single base pair change can be highly problematic. It may code for a different amino acid and can result in a protein whose behavior is dramatically different. This is especially true for PCa. It is worth a brief examination of the impact of SNPs as compared to CNVs. SNPs are the in extremis of a CNV, namely a single base pair. From Mikropoulos et al the authors note regarding PCa:

Several SNPs associated with PrCa risk in the 8q24 locus were among the earliest identified. The 8q24 region is a gene-poor region located upstream of the MYC proto-oncogene and this suggested an association with its expression, which was later proven to occur in a tissue-specific manner. Another important SNP is rs10993994 in the region containing the MSMB gene on chromosome 10. This risk allele associates with reduced MSMB protein expression. MSMB expression is high in normal and benign prostate tissue and low in PrCa. MSMB regulates cell growth and when lost, tumor cells grow in an uncontrolled manner. The odds ratio (OR) for this SNP's association to PrCa was established as 1.61. This is a potential biomarkersince urine MSMB assays have been developed and their role in screening is being evaluated.

SNP rs2735839 was identified between the KLK2 and KLK3 genes on chromosome 19 where there is a kallikrein gene cluster. Kallikreins are serum proteases and the most well-known member of this group is the prostate-specific antigen (PSA), which is widely used for screening and monitoring PrCa. SNPs were also identified in the intronic region areas of the LMTK2 gene, which codes for cdk5, the SLC22A3 gene, which codes for an organic cation transporter and NUDT10, which regulates DNA phosphorylation.

It is estimated that over 1800 SNPs may be associated with PrCa risk. *MacInnis et al. described, in 2011, a risk prediction algorithm based on common genetic variants. Based on segregation analysis from 4390 families with significant history of PrCa they created a model of inheritance that included a polygenic (P) component using 26 PrCa susceptibility SNPs; the P-model. Zheng et al. evaluated 16 SNPs in 2893 PrCa patients and 1781 controls in a Swedish population. The mathematical model combining the five SNPs with patients' family history provided an OR of 9.46 for predicting PrCa.*

The statement of there being over 1800 SNPs alone being related to PCa demonstrates the fragility of many of the coding genes. Waddell et al have applied SNP analysis to multiple myeloma and their approach has some significance here as well. They do also provide insight into problems with SNPs:

Despite these advantages, SNP data does present three major challenges for our approach. The first challenge of SNP data is that there are now well over 1.8 million SNPs known, but measuring them all is typically cost-prohibitive. Hence, in contrast to microarray data where measurements are recorded for a substantial fraction of the known genes, SNP data contains measurements for only a small fraction of the known SNPs – typically a few thousand. Therefore, it is quite possible that, for a given classification task, the features that would allow for highly accurate prediction will be missing.

Second, missing values are more common in SNP data than in microarray data. This must be taken into consideration when choosing a learning algorithm, since some methods are more capable of handling missing data than others. Third and perhaps most interesting, SNP data is "unphased." ... The result of SNP data being unphased is that this additional, and potentially highly informative, phase information is not available for model building.

Thus the same issues regarding CNVs are present in SNPs. Thus perhaps a more detailed analysis of both along with miRNAs and even methylations are important. The issue is also; what is the process and the progression of these changes? What causes them? Both CNV and SNPs seem to be the result of defective mitotic replication, but what is the reason and process associated with that? One may further ask then what can be changed in such a process and is there a potential path for a therapeutic?

4 SOME RESULTS

The main issue in the Lancet paper was to create groups based on certain CNV profiles that were prognostic of outcome. Returning to the Lancet study, we now present a summary of their results. They found:

Unbiased hierarchical clustering in the Toronto cohort (appendix pp 45–46) showed four localised prostate cancer subtypes with distinct genomic profiles:

1. *subtype 1 (characterised by gain of chromosome 7),*
2. *subtype 2 (deletion of 8p and gain of 8q),*
3. *subtype 3 (loss of 8p and 16q), and*
4. *subtype 4 (so-called quiet genomes due to few genomic alterations).*

Subtypes 2 and 3 share many common genetic alterations (504 genes altered in >25% of patients in both subtypes), but χ^2 tests showed eight regions that differed significantly, including gain of 8q in subtype 2 and 16q deletion in subtype 3 (appendix p 14). All four subtypes were confirmed in the MSKCC radical prostatectomy cohort and were not associated with TMPRSS2–ERG fusion, Gleason score, or T category (appendix pp 15–17, 47–48).

We summarize their results regarding specific CNVs in a Table derived from the Lancet paper shown below:

Target	*Type*	*Rank 1*	*Rank 2*	*Genes in Region*
8p21.3	Del	1	5	*PEBP4, RHOBTB2, TNFRSF10B, TNFRSF10C, TNFRSF10D,TNFRSF10A, CHMP7, LOXL2, ENTPD4*
8p11.22	Del	42	1	*FGFR1, C8orf86*
8p23.1	Del	2	2	*DEFB103A, DEFB103B, SPAG11B, DEFB104A, DEFB104B, DEFB106A, DEFB106B, DEFB105A, DEFB105B, DEFB107A, DEFB107B, SPAG11A, DEFB4*
8p22.1	Del	3	3	***NKX3-1**, STC1*
8q24.3	Amp	29	78	*COL22A1 KCNK9 TRAPPC9 CHRAC1 EIF2C2 PTK2 DENND3 SLC45A4 GPR20 PTP4A3 FLJ43860 TSNARE1 BAI1 ARC JRK PSCA LY6K C8orf55 SLURP1 LYPD2 LYNX1 LY6D GML*
8q21.2	Amp	7	167	*REXO1L1*
16q22.2	Del	16	9	*HP, HPR, TXNL4B, DHX38, PMFBP1, ZFHX3*
16q23.2	Del	6	52	*WWOX, MAF, DYNLRB2, CDYL2, C16orf61, CENPN, ATMIN, C16orf46, GCSH, PKD1L2, BCMO1, GAN, CMIP*
6q15	Del	13	17	*MAP3K7, BACH2*
5q11	Del	16	60	*LRCH1 ESD HTR2A SUCLA2 NUDT15 MED4 ITM2B RB1 P2RY5 RCBTB2 CYSLTR2 FNDC3A MLNR CDADC1 CAB39L SETDB2 PHF11 RCBTB1 ARL11 EBPL KPNA3 C13orf1 TRIM13 KCNRG*

The Lancet paper continues as follows:

Patients classified as subtype 4 had a significantly better prognosis compared with those with other subtypes …. Biochemical relapse-free survival at 5 years was:

- *58% (95% CI 37–92) for subtype 1,*
- *55% (37–81) for subtype 2,*
- *53% (37–78) for subtype 3, and*
- *89% (84–94) for subtype 4.*

Subtype 1 seems to be characterised by increased relapse after 3 years rather than increased risk at all times, but larger cohorts are needed to clarify this finding. These subtypes are prognostic for biochemical recurrence by 18 months …, which is associated with increased prostate-cancer-specific mortality. Indeed, in the Toronto cohort, being subtype 2 was associated with significantly worse overall survival than being in subtype 4

The subtype 4 which has no significant CNVs had the best prognosis. Yet it was 89% survival and thus there clearly were other factors leading to mortality. Subtypes 2 and 3 seem the most sensitive to CNV regarding survival and type 1 is not far behind.

The issues of ongoing concern remain; (i) what is the cause of these CNVs, and (ii) if we know that the outcome is poorer than what can we do for the patient? As of this time both do not seem to have adequate answers.

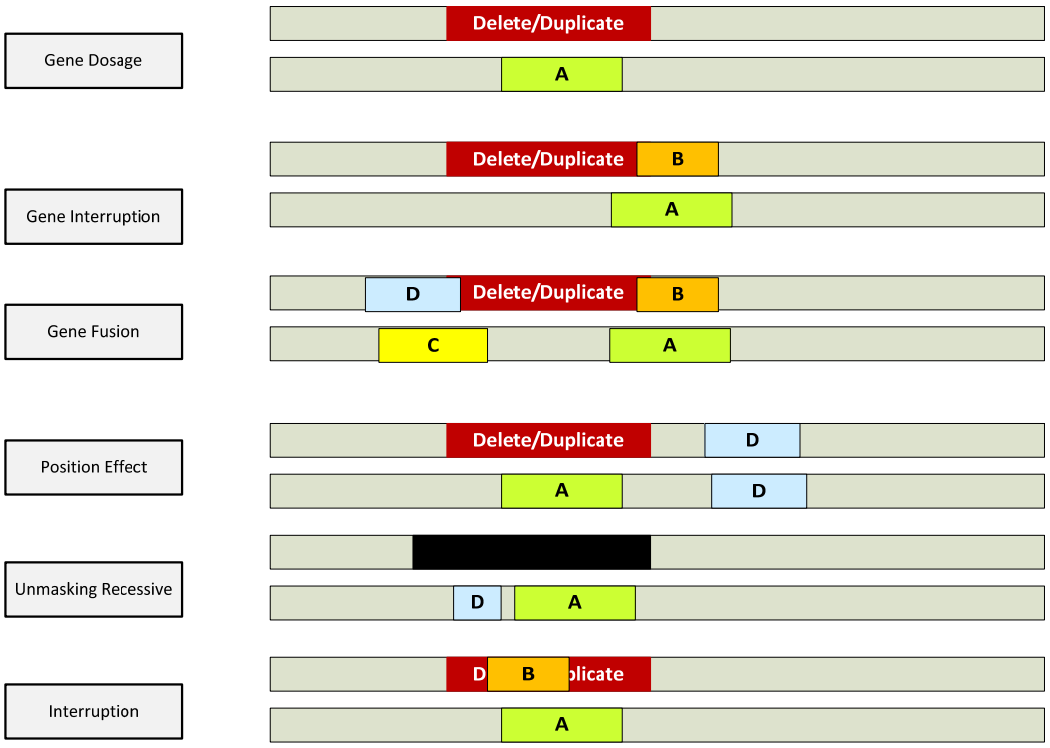
5 TECHNIQUES

We now want to discuss the method used in the analysis because it has substantial merit. We first summarize what effects gene changes can have and provide a high level comparison. We then details a bit more on CNVs and then we discuss the technique.

5.1 GENE POSITION AND COUNT EFFECTS

One question we may ask is; what does a CNV do to gene expression? Clearly if the extra segment falls on top of the existing gene or a deletion occurs at the same place we have an obvious effect. But what if we duplicate then gene? Simply several things may happen. We may have increased expression resulting in possible aberrant behavior. We may also have lost a promoter and actually lose expression.

From Lupiski and Shaffer we have the following Figure (modified) and it is described below:



The authors then comment on the above as follows:

Six models are depicted and include

(A) gene dosage, where there is a dosage sensitive gene within the rearrangement;

(B) gene interruption, wherein the rearrangement breakpoint interrupts a gene;

(C) gene fusion whereby a fusion gene is created at the breakpoint that either fuses coding sequences or a novel regulatory sequence to the gene;

(D) position effect, in which the rearrangement has effects on expression/ regulation of a gene near the breakpoint, potentially by removing or altering a regulatory sequence;

(E) unmasking recessive allele, where a deletion results in hemizygous expression of a recessive mutation or further uncovers/exacerbates effects of a functional polymorphism; and

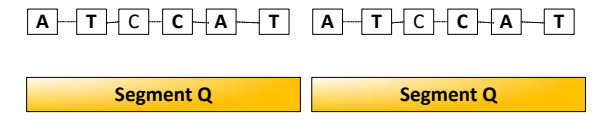
(F) by potentially interrupting effects of transvection, where the deletion of a gene and its surrounding regulatory sequences affects the communication between alleles.

Thus when we consider CNVs and their effects we may very well have some amalgam of all of the above elements. In a sense we may see similar effects in methylation on cells.

5.2 COPY NUMBER VARIATIONS (CNV)

Copy Number Variations are an amalgam of insertions, deletions, duplications and overwrites. They also may include translocations and some fusing effects. They are quite common in both intron and exon regions and are for the most part benign. They can be inherited and also found de novo in cells. They are frequent in many cancer cells and if one were to compare a benign somatic cell to say a malignant prostate cancer cell then one may likely find many CNVs which occur in regions of the genome which are locations of control genes. The result is one of the many gene modifying modes we had examined above.

Now we review the CNV structure as below:



CNV Duplication Gene segment of 3 genes shown.

Regular Gene segment of 3 genes shown.

This is a simple case. It represents a modest duplication of a gene. However not all such CNVs are so simple? In fact they are quite complex and tend to mask portions of expressed or controlled genes as well as regions which take an active part in expression such as promoter regions.

Now there are several questions we need to examine in looking at CNVs in PCa. Some are:

1. What is the cause of a CNV? What happens and when does it happen. We know that malignant cells have increased mitotic behavior and it is in the duplicating of the DNA that this most likely happens. But perhaps we also have an occurrence during the reading of the genes in what would be a normal process. Thus far the only reasonable guess is some disturbance during mitosis.

2. Is there a therapeutic target for preventing or reversing this process?

3. Where does the process begin? Is it while the cells are still in the prostate and can we correlate Gleason grade to this as well?

4. Is there a mechanism for CNV creation dependent on preceding genetic aberrations, such as a simple methylation suppressing a promoter region? Namely can inflammation be a precipitating event?

The list of questions can become significant. These of course are important as we try to construct models for cancer progression. All too often the modelers create systems reliant upon genetic activation or suppression only and fail to account for what may be considered secondary effects but which are prime movers in the process. We shall return to this in a later report.

5.3 ARRAY COMPARATIVE GENOMIC HYBRIDIZATION (ACGH)

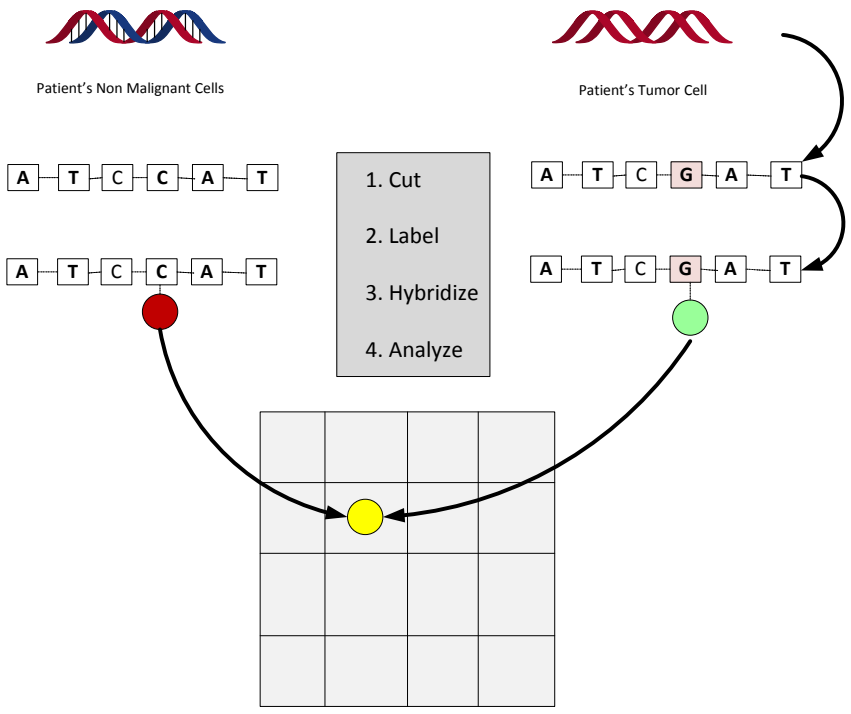
One of the techniques used to identify CNVs is array Comparative Genomic Hybridization, aCGH, a technique that uses microarray chips along with the comparison of tagged samples of DNA from a standard and a sample to be tested. It is a relatively elegant and simple method and has the potential for high throughput. As Bejjani and Shaffer state:

Microarray-based comparative genomic hybridization (array CGH) is a revolutionary platform that was recently adopted in the clinical laboratory. This technology was first developed as a research tool for the investigation of genomic alterations in cancer. It allows for a high-resolution evaluation of DNA copy number alterations associated with chromosome abnormalities. Array CGH is based on the use of differentially labeled test and reference genomic DNA samples that are simultaneously hybridized to DNA targets arrayed on a glass slide or other solid platform.

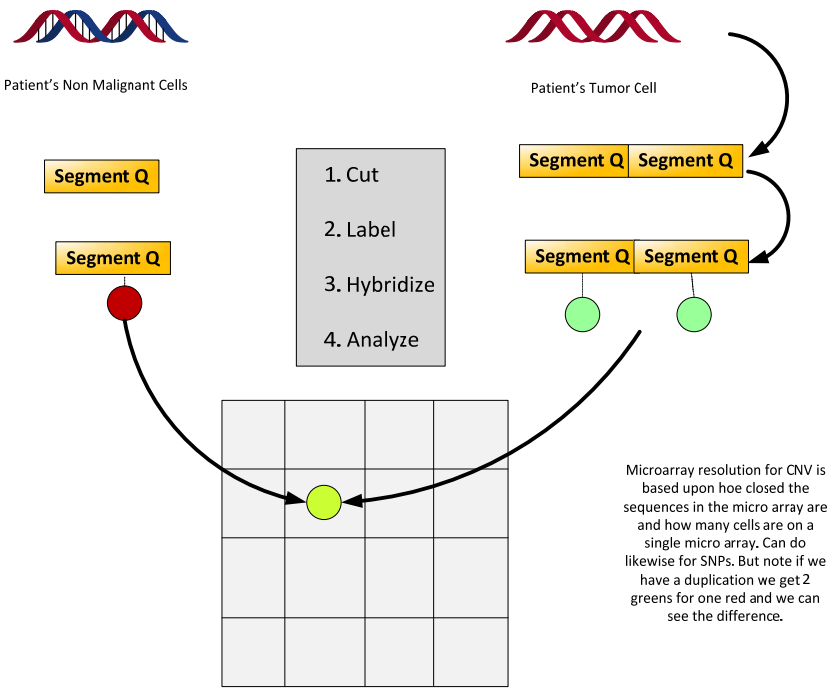
In this review, we examine the technology and its transformation from a research tool into a maturing diagnostic instrument. We also evaluate the various approaches that have shaped the current platforms that are used for clinical applications. Finally, we discuss the advantages and shortcomings of "whole-genome" arrays and compare their diagnostic use to "targeted" arrays.

Depending on their design, microarrays provide distinct advantages over conventional cytogenetic analysis because they have the potential to detect the majority of microscopic and submicroscopic chromosomal abnormalities. This new platform is poised to revolutionize modern cytogenetic diagnostics and to provide clinicians with a powerful tool to use in their increasingly sophisticated diagnostic capabilities.

Looking at DNA, cDNA or even mRNA, we often break it up by use of certain enzymes, and then tag the elements and then use a microarray to "stick" them in pre-marked cells. We demonstrate this below in a simplified graphic.



Now we consider applying this to a CMV. Note that a CMV is a long segment of a genome and we can select then and reproduce them but if there are two or more we can see they have a stronger influence and are just not neutralized. We show a simplified example below.



Thus this elegant technique blends the use of florescent tagging of test and target segments of DNA in a microarray system. The output can be plotted in some genomic sequence following the colors. Variations of positive and negative values can show the excess or depletion of gene segments depending on which colors are detected. This in a very rapid manner a whole genome can be scanned for CNVs and determinations made as to where they are present.

From Bejjani and Shaffer we have:

Array CGH is based on the same principle as traditional metaphase CGH. In both techniques, whole genomic DNA from a control (or reference) and genomic DNA from a test (or patient) are differentially labeled with two different fluorophores and used as probes that are cohybridized competitively onto nucleic acid targets. In traditional metaphase CGH, the target is a reference metaphase spread.

In array CGH, these targets can be oligonucleotides, cDNAs, or genomic fragments that are cloned in a variety of vectors such as plasmids, cosmids, BACs, or P1 artificial chromosomes. In this review, we will restrict our discussion to array CGH that uses BACs as hybridization targets because oligonucleotide arrays and cDNA arrays are not currently used in clinical diagnostics.

The resolution of array CGH is defined by two main factors:

1) the size of the nucleic acid targets and

2) the density of coverage over the genome; the smaller the size of the nucleic acid targets and the more contiguous the targets on the native chromosome, the higher the resolution of the array.

Furthermore, a comparison of ratios between overlapping clones can narrow the region of copy-number change to within a fraction of a clone length because the fluorescence ratio for each clone represents the average copy-number ratio over the length of the entire clone.18 The sensitivity and quantitative potential of array CGH for gene dosage measurements has been reviewed, and the usefulness of this technique in identifying gene copy number abnormalities associated with cancer has been demonstrated.

From a modified version of Mikhail we have a comparison of the two techniques[10]:

Array Comparative Genomic Hybridization (Array CGH)	***SNP arrays***
Single-sequence oligonucleotides of ~60 bp	Two 20–60 bp oligonucleotides of different sequence
Two labeled DNAs (patient and control) per hybridization	Only patient DNA labeled and hybridized
Resolution down to size of oligonucleotides; exon by exon	Resolution limited by SNP distribution
No detection of UPD or consanguinity	Able to detect consanguinity and most UPD
Limited SNP addition possible recently	Detection of most known clinically relevant CNVs but not exon by exon

We have shown the above Table to more clearly demonstrate the differences between CNV and SNPs.

[10] http://www.uab.edu/hcgs/images/PDF_documents/Genotyping_and_CNV_IV.pdf This is a good set of charts with an overview. The only rather shabby element is the introduction of their politically "correct" slides demeaning a US President. Regrettably this is an all too common childish and unprofessional characteristics of many academics.

6 RELATED GENES

We present a brief summary of the related genes discussed. We believe it is worth an inclusion of these to refresh what the specific genes the Lancet article refers to. Of interest is StAR or STAR which regulates cholesterol. One should note that cholesterol and testosterone, an androgen, are in the same secondary pathways and androgen receptors are key players in PCa metastasis.

6.1 STAR

The steroidogenic acute regulatory protein (StAR) is produced by the STAR gene. It regulates steroid production and appears to have weak linkages to prostate cancer.

From Brinkmann we have the following:[11]

STAR: ANDROGEN BIOSYNTHESIS Androgens (testosterone and 5α-dihydrotestosterone) belong to the group of steroid hormones. The major circulating androgen is testosterone, which is synthesized from cholesterol in the Leydig cells in the testis. Testosterone production in the fetal human testis starts during the sixth week of pregnancy. Leydig cell differentiation and the initial early testosterone biosynthesis in the fetal testis are independent of luteinizing hormone (LH).

During testis development production of testosterone comes under the control of LH which is produced by the pituitary gland. Synthesis and release of LH is under control of the hypothalamus through gonadotropin-releasing hormone (GnRH) and inhibited by testosterone via a negative feedback mechanism.

The biosynthetic conversion of cholesterol to testosterone involves several discrete steps, of which the first one includes the transfer of cholesterol from the outer to the inner mitochondrial membrane by the steroidogenic acute regulatory protein (StAR) and the subsequent side chain cleavage of cholesterol by the enzyme P450scc. This conversion, resulting in the synthesis of pregnenolone, is the rate-limiting step in testosterone biosynthesis. Subsequent steps require several enzymes including, 3β-hydroxysteroid dehydrogenase, 17α-hydroxylase/C17-20-lyase and 17β-hydroxysteroid dehydrogenase type 3

From NCBI we have:

The protein encoded by this gene plays a key role in the acute regulation of steroid hormone synthesis by enhancing the conversion of cholesterol into pregnenolone. This protein permits the cleavage of cholesterol into pregnenolone by mediating the transport of cholesterol from the outer mitochondrial membrane to the inner mitochondrial membrane. Mutations in this gene are

[11] ANDROGEN PHYSIOLOGY: RECEPTOR AND METABOLIC DISORDERS, by Albert O. Brinkmann, Ph.D. Assoc. Professor of Biochemistry, Endocrinology, and Reproduction, Department of Reproduction and Development , Erasmus MC, University Medical Center Rotterdam, PO Box 1738, 3000 DR Rotterdam, Netherlands, Revised November, 2009

a cause of congenital lipoid adrenal hyperplasia (CLAH), also called lipoid CAH. A pseudogene of this gene is located on chromosome 13.

Thus as part of the above pathways it is reasonable to see it part of the androgen control mechanism.

As Kocerha states:

Since a direct link between β-sitosterol exposure and TGF-β upregulation in prostate cancer cells has been shown ,… Interestingly, a study showed TGF-β protein levels are upregulated by at least 50% in prostate stromal cells after a 6 day exposure to β-sitosterol, one major chemical found in paper mill contaminants

She also states:

Researchers had known for many years that de novo protein synthesis was required for steroid synthesis; however, the identity of the protein involved in facilitating the transport of cholesterol across the mitochondria eluded them until 1995. A 30 kDa protein in rat adrenal cells stimulated by ACTH, now known as StAR, was first observed when the [35S] methionine labeled proteins were electrophoresed through a 2D gel. … StAR is a mitochondrial protein synthesized as a 37 kDa precursor protein in the cytosol of mammalian cells.

Upon stimulation, the 37 kDa precursor is targeted via its signal sequence to the mitochondria. As the precursor protein is imported into the mitochondrial inner compartment, the protein's signal sequence is removed by a matrix processing protease and contact sites are formed between the outer and inner mitochondrial membranes. The precursor protein is further processed by the mitochondrial intermediate processing peptide to remove the targeting sequence, forming the mature 30 kDa protein. The cytoplasmic 37 kDa protein has a half-life of around 10-15 minutes; however, the 30 kDa inactive protein has a longer half-life of up to several hours.

6.2 NKX3-1

NKX3-1 (also NKC3.1) is a gene and Nkx3.1 its protein which is putatively a tumor suppressor gene which is primarily prostate specific. As c-Myc tends to regulate the transcription of many genes, Nkx3.1 regulates the control mechanism for the prostate cells. Even more specifically it has been argued that Nkx3.1 regulates the luminal cell growth by Iwata et al. Yet Goldstein et al would argue from similar murine models that use of up-regulated ERG translocations with Akt activation, namely putatively suppressing PTEN, allows for basal cell growth rather than luminal cell transformation.

It has been observed that diminished expression of NKX3.1 (8p21) is associated with prostate cancer progression in humans, and in mice, loss of nkx3.1 leads to epithelial cell proliferation and altered gene expression patterns….Loss of heterozygosity of 8p21 is observed in a high percentage of intraepithelial prostatic neoplasia and early carcinoma lesions, strongly implicating this region in the initial stages of prostate carcinogenesis. The importance of NKX3.1 as a dose-

dependent regulator of prostate epithelial cell growth is strongly supported by analyses of nkx3.1 knockout mice. Homozygous nkx3.1 mutant mice develop prostate epithelial hyperplasia and dysplasia that progresses with age, and lesions with histologic features strongly resembling human prostatic intraepithelial neoplasia develop in homozygous mice between 1 and 2 years of age. Importantly, both hyperplasia and prostatic intraepithelial neoplasia-like lesions also occur in a significant proportion of nkx3.1 heterozygous mutants

The question then is, is NKX3.1 a true tumor suppressor gene? Despite that loss of function of Nkx3.1 predisposes to prostate cancer, it is not sufficient for tumorigenesis.(Shen 2003) Moreover, while one allele of NKX3.1 is lost by means of chromosomal deletion in PIN and prostate cancer, the other allele does not undergo mutational inactivation, although protein expression is epigenetically down-regulated or lost. These features, along with the relatively subtle consequences following forced expression of Nkx3.1 in prostate cancer cells are not consistent with activities of "classic" tumor suppressor genes, such as p53, Rb, or Pten. Instead, Nkx3.1 appears to act more like a tumor modulator, serving as a regulator of differentiation, which in turn prevents cancer initiation. In this regard, further analyses of NKX3.1 can provide important insights into the relationship between regulation of differentiation and carcinogenesis.

As Iwata et al have observed:

The prevailing model of NKX3.1 expression in human prostate cancer suggests that while the protein may decrease in PIN lesions, it is much more commonly decreased in invasive adenocarcinomas, and nearly completely lost in most, if not all, metastatic prostate adenocarcinomas … there was a variable decrease in expression of Nkx3.1 in PIN lesions, and that Nkx3.1 was virtually completely lost in invasive adenocarcinomas… several observations from our group regarding NKX3.1 differ from this prevailing view. First, in a previous report, while reductions of NKX3.1 protein occurred in PIN lesions and some adenocarcinomas, the reductions were relatively minor and virtually all invasive adenocarcinomas retained significant levels of NKX3.1 protein …

More recently we have found that the majority of very high grade (Gleason score 8–10) localized prostate cancers … retain high levels of expression of NKX3.1 protein. In the present study we found that, as compared to high grade PIN, the staining for Nkx3.1 protein actually increased substantially in pre-invasive cribriform PIN/CIS lesions and in early invasive adenocarcinomas, and these levels correlated inversely with levels of MYC expression.

These results indicate the Nkx3.1 may be dynamically regulated during progression of this disease. … It is possible, therefore, that Nkx3.1 expression in invasive prostatic acini in MYC-driven mouse prostate cancers may represent a recapitulation or caricature of the process of stromal invasion/branching morphogenesis in development, and, that Nkx3.1 may facilitate this process. …

Lei et al. found that forced restoration of Nkx3.1 expression in Pten null epithelium led to decreased cell proliferation, increased cell death, and prevention of tumor initiation.. They further showed that Nkx3.1 was required to engage the p53 pathway, indicating that reduced Nkx3.1 expression can itself abrogate p53 signaling.

These findings raise the interesting possibility that the reduction in Nkx3.1 seen upon the induction of MYC in the mouse prostate prevents the induction of p53 induced apoptosis, thus facilitating MYC's ability to transform these cells. Additional studies in which Nkx3.1 expression is kept at high levels during induction of MYC in prostate epithelium will be required to address this question further. We do not know precisely how MYC is regulating Nkx 3.1 protein expression, …"

Specifically Iwata et al state:

Since MYC may downregulate Nkx3.1 at the level of transcription …, it is possible that elevated MYC itself may be responsible for down-regulating Nkx3.1 expression.

In effect, this implies that MYC controls NKX3.1 and thus up-regulated MYC results in a down regulated NKX3.1. If NKX3.1 is controlling prostate stability then its overall regulation is via MYC. Controlling and suppressing MYC would control and up-regulate NKX3.1 and thus stabilize prostate growth. The complete pathway for this gene does not seem to be complete at this stage. Its importance is well defined however.

From Eide et al:

In this study we show both that TWIST1 mRNA is up-regulated by androgen via AR and that NKX3-1, a well-known androgen-regulated gene, binds the upstream regulatory region of the TWIST1 gene and represses the expression of TWIST1… The physical binding of NKX3-1 to the TWIST1 promoter might block the mesenchymal drive of TWIST1, until NKX3-1 expression is down-regulated or lost in PIN or adenocarcinoma lesions. Loss of NKX3-1 expression has been observed in ~20% of PIN lesions, ~40% of advanced prostate tumors and up to 80% of metastatic prostate cancer.

Androgen deprivation therapy as the most widely used treatment for advanced prostate cancer is likely to abolish androgen-stimulation of NKX3-1, leading eventually to down-regulation of repressor protein and de-repression of TWIST1's metastatic potential. In an attempt to identify genes whose regulation are altered by NKX3-1, Song et al. performed gene expression profiling analyses on micro dissected glands from NKX3-1- deficient prostate tissues during prostate cancer progression. They observed similarities between the expression profile of the micro dissected glands and constitutive activated AKT transgenic mice as well as PTEN-deficient mice, suggesting that the PTEN-AKT-NKX3-1 axis serve as a major molecular path of prostate tumorigenesis.

Li and Zhou showed that activation of the AKT pathway by TWIST1 is critical for the sustention of cancer stem cell like traits generated by EMT, again suggesting a link between loss of NKX3-1 expression, relive of TWIST1 expression and eventually activation of AKT pathway. Conclusions We report in this paper that TWIST1 is an androgen-regulated gene, tightly regulated by NKX3-1. We show that NKX3-1 binds to the TWIST1 promoter and that NKX3-1 overexpression reduces the activity of a TWIST1 promoter reporter construct, whereas NKX3-1 siRNA up-regulated endogenous TWIST1 mRNA in prostate cancer cells. Our finding that NKX3-1

represses TWIST1 expression emphasizes the functional importance of NKX3-1 in regulating TWIST1 expression during prostate cancer progression to metastatic disease.

6.3 MYC

Myc or specifically c-Myc, is a powerful gene element which induces cell growth. c-Myc is so strong promoter of cell proliferation and growth. c-Myc is a transcription factor which is essential in the growth and expansion of the cell.

In the paper by Iwata et al the authors examine its influence during the development of PIN. They state:

Lo-MYC and Hi-MYC mice develop prostatic intraepithelial neoplasia (PIN) and prostatic adenocarcinoma as a result of MYC overexpression in the mouse prostate[1]. However, prior studies have not determined precisely when, and in which cell types, MYC is induced. Using immunohistochemistry (IHC) to localize MYC expression in Lo-MYC transgenic mice, we show that morphological and molecular alterations characteristic of high grade PIN arise in luminal epithelial cells as soon as MYC overexpression is detected.

These changes include increased nuclear and nucleolar size and large scale chromatin remodeling. Mouse PIN cells retained a columnar architecture and abundant cytoplasm and appeared as either a single layer of neoplastic cells or as pseudo-stratified/multilayered structures with open glandular lumina—features highly analogous to human high grade PIN.

Also using IHC, we show that the onset of MYC overexpression and PIN development coincided precisely with decreased expression of the homeodomain transcription factor and tumor suppressor, Nkx3.1. Virtually all normal appearing prostate luminal cells expressed high levels of Nkx3.1, but all cells expressing MYC in PIN lesions showed marked reductions in Nkx3.1, implicating MYC as a key factor that represses Nkx3.1 in PIN lesions.

To determine the effects of less pronounced overexpression of MYC we generated a new line of mice expressing MYC in the prostate under the transcriptional control of the mouse Nkx3.1 control region. These ''Super-Lo-MYC'' mice also developed PIN, albeit a less aggressive form. We also identified a histologically defined intermediate step in the progression of mouse PIN into invasive adenocarcinoma. These lesions are characterized by a loss of cell polarity, multi-layering, and cribriform formation, and by a ''paradoxical'' increase in Nkx3.1 protein. Similar histopathological changes occurred in Hi-MYC mice, albeit with accelerated kinetics.

Our results using IHC provide novel insights that support the contention that MYC overexpression is sufficient to transform prostate luminal epithelial cells into PIN cells in vivo. We also identified a novel histopathologically identifiable intermediate step prior to invasion that should facilitate studies of molecular pathway alterations occurring during early progression of prostatic adenocarcinomas.

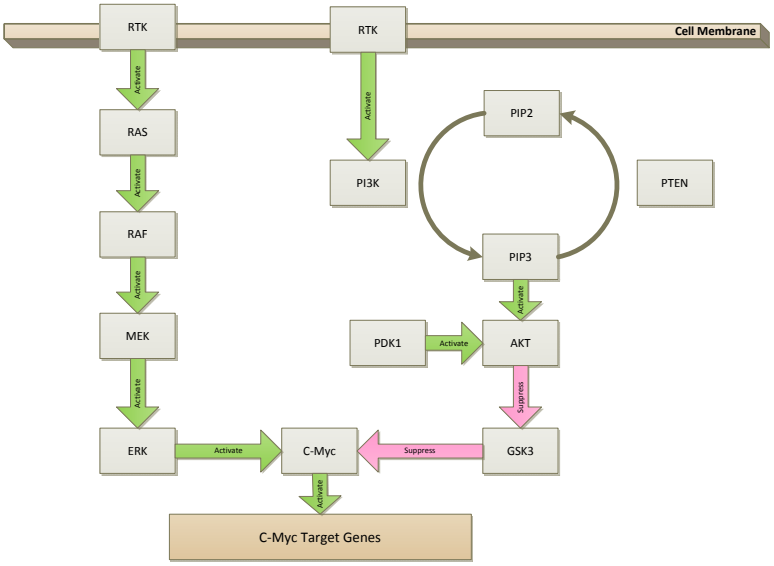
In the following graphic we depict the influence elements on c-Myc. This is a complex system of interlinking genes which when expressed in the correct manner can slow cell over expansion.

The chart below is a modification from Bunz (p. 203) and it shows the gross characteristics of this control path. PTEN is a key element in control. What this does not show are two key elements, and indirectly a third.

First it does not show the fact that these are protein concentrations at work, one influencing the other and so forth. There is a feedback mechanism missing.

Second, it does not portray the temporal elements, namely this is a static gross representation of the influencing factors as if done in some generic snapshot. I fact the concentrations are time varying and it is this time variation which when combined with the feedback loops renders certain instabilities leading to malignancy, namely uncontrolled growth.

Third, it fails to show the other genes and specifically the feedback mechanism of these genes. Namely PTEN is influence by these.



As Deutsch et al state:

MYC is one of the earliest genes to be expressed in the prostate gland after androgen stimulation. Transgenic mice that overexpress Myc develop prostate neoplasia. The importance of MYC overexpression in prostate cancer has been corroborated by the discovery of an increased copy number of the portion of 8q containing MYC in prostate-cancer samples, and particularly in advanced cases.

Only 6% of primary tumours show a gain of 8q sequences, whereas this change is seen in 89% of tumor recurrences after hormonal therapy.
MYC (a downstream target of AR) has been shown to be regulated by the AR and to be needed for AR-dependent and AR-independent growth. These data suggest that MYC might be involved in the development of androgen-independent prostate cancer, including that resulting from an increase in AR signalling.

From Iwata et al:

….we show that the onset of MYC overexpression and PIN development coincided precisely with decreased expression of the homeodomain transcription factor and tumor suppressor, Nkx3.1. Virtually all normal appearing prostate luminal cells expressed high levels of Nkx3.1, but all cells expressing MYC in PIN lesions showed marked reductions in Nkx3.1, implicating MYC as a key factor that represses Nkx3.1 in PIN lesions. To determine the effects of less pronounced overexpression of MYC we generated a new line of mice expressing MYC in the prostate under the transcriptional control of the mouse Nkx3.1 control region.

6.4 PTEN

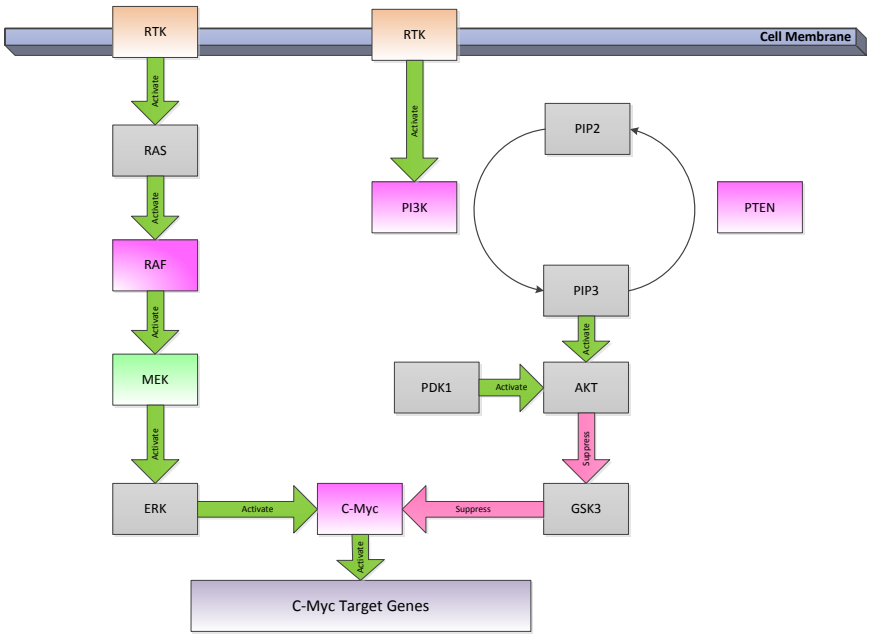
PTEN has been one of the PCa specific genes whose loss of function is directly related to PCa. We have examined this in detail in prior works.

PTEN is a significant gene which controls the Akt pathway which in turn controls the replication of cells. Loss of PTEN is often seen in metastatic prostate cancer. In many ways it is the hallmark of this change. As stated in NCBI[12]:

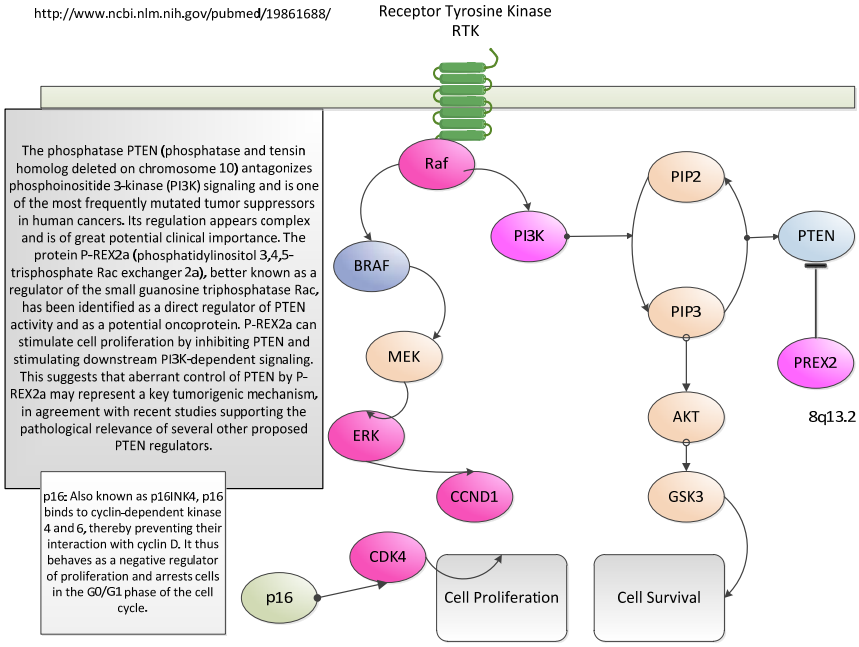
This gene was identified as a tumor suppressor that is mutated in a large number of cancers at high frequency. The protein encoded this gene is a phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase. It contains a tensin like domain as well as a catalytic domain similar to that of the dual specificity protein tyrosine phosphatases. Unlike most of the protein tyrosine phosphatases, this protein preferentially dephosphorylates phosphoinositide substrates. It negatively regulates intracellular levels of phosphatidylinositol-3,4,5-trisphosphate in cells and functions as a tumor suppressor by negatively regulating AKT/PKB signaling pathway.

First the PTEN pathway as shown below:

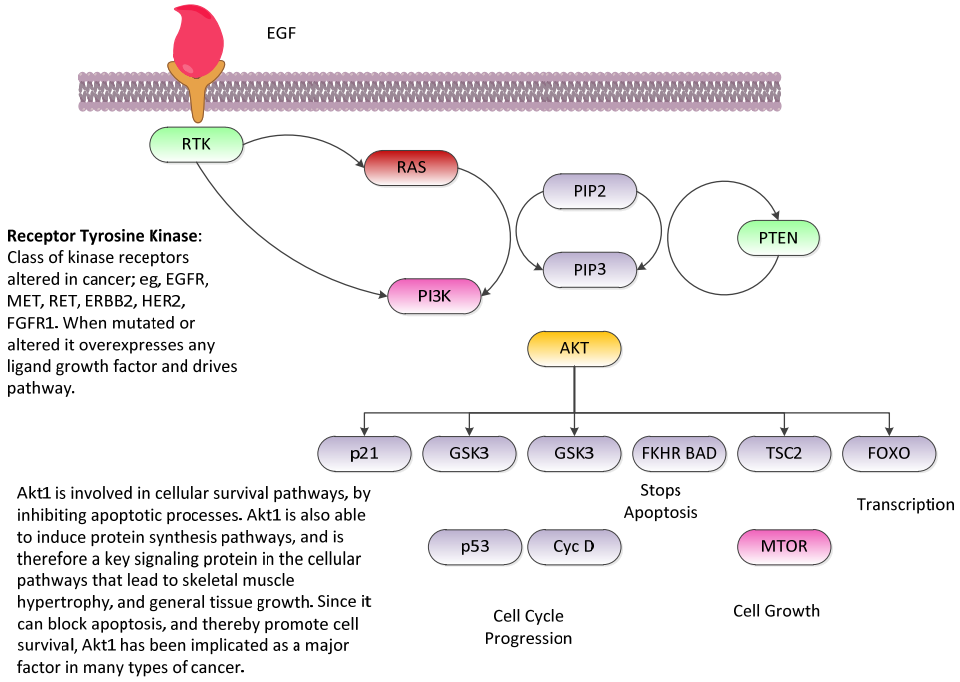
[12] http://www.ncbi.nlm.nih.gov/gene/5728



Note PTEN modulates the production of Akt which in turn modulates c-Myc which in turn controls cell reproduction. Any effect which causes PTEN not to be expressed will, in turn, results in unfettered cell growth.



PTEN has become a key gene in the development of prostate cancer. It controls a pathway leading up to c-myc control and once PTEN is lost the PCa can be considered as very aggressive. Its loss results in an activation of Akt and then c-myc causing uncontrolled cell growth. The pathway is shown below:



As Jelovac and Park state[13]:

The phosphatase and tensin homolog gene (PTEN) is a tumor suppressor located on the human chromosome 10q arm and is an important mediator of carcinogenesis in a variety of human malignancies. By the strictest definition, a tumor suppressor is a gene whose loss confers an increased lifetime risk of developing tumors. The most illustrative examples of genes that fulfill this criterion are those associated with familial cancer syndromes whereby heritable inactivation of 1 allele and subsequent increased tumor risk is passed along to each generation in an autosomal-dominant fashion.

Using this as a framework, PTEN is a bona fide tumor suppressor gene in that heritable germline mutations have been described in Cowden syndrome (CS), giving rise to a number of human tumors and cancers, most notably thyroid and breast cancers. As is the paradigm of tumor suppressor genes, affected patients with CS inherit 1 mutant inactive copy of PTEN from either parent, and the ensuing loss of the second allele results in tumor formation with subsequent genetic events that eventually lead to cancer. Although there are notable exceptions to this model, most heritable cancer syndromes are believed to adhere to this pattern.

From the work of McMenamin et al we have the slides below. Here is a case where PIN is still expressing PTEN but as we increase the grade of PCa we see the elimination of PTEN expression. Thus we can say that PIN is a state prior to PTEN suppression and a corollary may be that PCa aggressiveness is reflective of loss of PTEN and activation of Akt pathway.

[13] http://jama.ama-assn.org/content/304/24/2744.full

7 OBSERVATIONS

One is always amazed at how a small bit of insight, perhaps well interpreted, may get presented by the Press as the be all and end all. It is especially true with cancer. The most recent case in point is a study by physicians at a Toronto Hospital as well as at MSKCC and other places that the existence of some CNV, copy number variants, can give some modest prognostic data on prostate cancer. Simply, using some 300 patients in toto, then compared tumor cells to non-tumor cells and determined what CNVs across the genome could possibly be prognostic. With the 300 patient sample they got an ROC with about a 70% AUC.

But now to the Press. A Canadian newspaper states[14]:

Canadian researchers have developed a genetic test to identify which men are at highest risk for recurrence of prostate cancer following localized treatment with surgery or radiation therapy. The genetic test provides a quick and highly accurate tool to determine which men with prostate cancer would do well with only surgery or radiation, and those who would need additional treatment — chemotherapy and hormone therapy, say the researchers, whose findings are described in Wednesday's online edition of the journal Lancet Oncology.

"Our findings set the stage to tackle the ongoing clinical problem of under-treating men with aggressive disease that will recur in 30 per cent to 50 per cent of patients due to hidden, microscopic disease that is already outside the prostate gland during initial treatment," said ...a clinician-scientist ... in Toronto. "This genetic test could increase cure rates in intermediate- to high-risk men by preventing progression to this metastatic spread of prostate cancer," said ... a scientist at the Ontario Institute for Cancer Research.

Now if one reads the opening sentence one could be led to believe that there is some massive discovery here. In my opinion, I will be delivering a more detailed analysis later, the following most likely are the results:

1. CNV are found everywhere in DNA. The fact that CNVs are more extensive in cancer cells may or may not be informative.

2. 300 patients is not a lot, especially not enough to in my opinion justify the first opening sentence. However in fairness the researcher did say "could".

3. There is the causative issue here. Why did these CNVs arise and why where they did. That seems to be the compelling issue.

One approach obviously is some breakdown in the control of DNA replication in mitosis. Another is the result of an increased inflammatory process. The review by Kundu and Surh is of interest in that it reviews the impact of inflammation and focuses on epigenetic factors of

[14] http://metronews.ca/news/canada/1210964/test-may-predict-prostate-cancer-recurrence-risk/

miRNA and cancer. However it can be argued that similar mechanisms may be interfering with the DNA replication as well.

In addition as Wojno et al note:

The diagnosis of prostate cancer is dependent on histologic confirmation in biopsy core tissues. The biopsy procedure is invasive, puts the patient at risk for complications, and is subject to significant sampling errors. An epigenetic test that uses methylation-specific polymerase chain reaction to determine the epigenetic status of the prostate cancer–associated genes GSTP1, APC, and RASSF1 has been clinically validated and is used in clinical practice to increase the negative predictive value in men with no history of prostate cancer compared with standard histopathology. Such information can help to avoid unnecessary repeat biopsies. The repeat biopsy rate may provide preliminary clinical utility evidence in relation to this assay's potential impact on the number of unnecessary repeat prostate biopsies performed in US urology practices.

4. Prognosis means that we can tell who after surgery or radiation will fare better or more poorly. But frankly so what if we have no way to mollify the negative results. We get to tell the patient that they are going to die at a greater odds ratio that someone else. We are not really certain but the odds are higher and yes we cannot do anything. Why even has that conversation?

This is a question about prognostic tests. Namely having them and obtaining a truly negative prognosis when there is no reasonable treatment is questionable. What do we tell the patient? Now we really know you will die, not at 50% but at 70%. Prognosis is of worth only if we can do something about it.

5. There are already dozens of similar tests, genes, mRNAs, CNV, SNPs proteins, exosomes, endosomes, and the list goes on. What does one get by utilizing this test?

We have examined dozens of proposed tests and most are based on genomic analysis of tissue from biopsies taken after surgery. It is known that a Gleason score of 7 or more portends a less favorable survival. The tests do not at this stage provide therapeutic guidance. We also continue to examine PSA post-surgery to measure extra prostatic recurrence.

Yet the real issue is why does the Press make it appear that there is truly something here, here? Let us address a few issues:

1. The analysis is broad and non-causal. It merely states that certain clusters of malignant cells have copy number variations of some sort and that is this appears at some broadly defined locations that there may be a higher incidence of an unfavorable outcome.

2. The major concern is always examining a cell by cell result. First one must be able to extract genetic material on a cell by cell basis. Then one must examine that content and compare it to the other such cells. Namely we would examine for stem cell presence.

3. How does this scheme fit with the now dozens of other such schemes.

4. Many epigenetic and epigenetic like processes are known to have oncogenic properties. Yet they do not appear as CNV characteristic.

5. What must be done to align the current prognostic measures based upon decades of clinical analysis with the results of this 300+ patient analysis. All too often the researchers seem to be overly optimistic and suggesting the introduction of a new test with quite limited clinical exposure.

6. The authors suggest that this test allows for selecting alternative treatments for those patients in a less favorable group. One must ask what treatments?

7. There may much too much exuberance in many CNV findings.

As McCarroll states:

The mapping of structural variation in our genomes has in five years traversed an arc that cartography traversed over centuries. What lessons might we draw for the exploration of other new continents, such as rare variants and epigenetics, in the search for the heritable basis of disease? One lesson is that new domains may turn out to be more familiar than they at first appear; many relationships may turn out to have been present in an overlooked form in earlier genome scale data sets.

Also, most discoveries may emerge after initial exuberance gives way to sober exploration. Insobriety about the unknown spurred human ancestors to explore new worlds. In science, a similar insobriety can lead us to mistake the nature of the novel, but it leads to the great efforts that ultimately get the work done. This human proclivity may yet be found to be an adaptive trait. The genetic mechanisms accounting for it—common or rare, structural or single nucleotide—are, at this point, anyone's guess.

As McCarroll notes many initial CNV studies aspired to results which were to be later tempered by limitations or prior observations using older techniques. There all too often is over-exuberance that must be tempered by further studies.

8 APPENDIX: DEFINITIONS

Definitions from Freeman et al:

1. Structural variant: A genomic alteration (e.g., a CNV, an inversion) that involves segments of DNA >1 kb.

2. Copy number variant (CNV); A duplication or deletion event involving >1 kb of DNA.

3. Duplicon :A duplicated genomic segment >1 kb in length with >90% similarity between copies

4. Indel: Variation from insertion or deletion event involving <1 kb of DNA.

5. Intermediate-sized structural variant (ISV): A structural variant that is ~8 kb to 40 kb in size. This can refer to a CNV or a balanced structural rearrangement (e.g., an inversion).

6. Low copy repeats (LCR): Similar to segmental duplication.

7. Multisite variant (MSV): Complex polymorphic variation that is neither a PSV nor a SNP.

8. Paralogous sequence variant (PSV): Sequence difference between duplicated copies (paralogs.)

9. Segmental duplication: Duplicated region ranging from 1 kb upward with a sequence identity of >90%. (Interchromosomal: Duplications distributed among nonhomologous chromosomes and Intrachromosomal: Duplications restricted to a single chromosome)

10. Single nucleotide polymorphism (SNP): Base substitution involving only a single nucleotide; ~10 million are thought to be present in the human genome at >1%, leading to an average of one SNP difference per 1250 bases between randomly chosen individuals

9 REFERENCES

1. Ahmed, H., Promoter Methylation in Prostate Cancer and its Application for the Early Detection of Prostate Cancer Using Serum and Urine Samples, Biomarkers in Cancer 2010:2.
2. Albertson, D., D. Pinkel, Genomic microarrays in human genetic disease and cancer, Human Molecular Genetics, 2003, Vol. 12, Review Issue 2.
3. Bejjani, B., L. Shaffer, Application of Array-Based Comparative Genomic Hybridization to Clinical Diagnostics, Journal of Molecular Diagnostics, Vol. 8, No. 5, November 2006
4. Brinkmann, A. O., ANDROGEN PHYSIOLOGY: RECEPTOR AND METABOLIC DISORDERS, Endocrinology, and Reproduction, Department of Reproduction and Development , Erasmus MC, University Medical Center Rotterdam, PO Box 1738, 3000 DR Rotterdam, Netherlands, Revised November, 2009.
5. Brooks, J., et al, CG Island Methylation Changes Near the GSTP1 Gene in Prostatic Intraepithelial Neoplasia, Cancer Epidemiology, Biomarkers & Prevention, "Col. 7, 531-536, June 1998.
6. Carroll, S., Copy number variation and human genome maps, Nature Genetics | volume 42 | number 5 | May 2010.
7. Carter, G., et al, PREDICTING THE EFFECTS OF COPY-NUMBER VARIATION IN DOUBLE AND TRIPLE MUTANT COMBINATIONS, Draft, supported by grants P50 GM076547 from NIGMS and by the ISB-University of Luxembourg Program.
8. Eide, T., et al, TWIST1, A novel androgen-regulated gene, is a target for NKX3-1 in prostate cancer cells, Cancer Cell International 2013, 13:4 doi:10.1186/1475-2867-13-4.
9. Freeman, J., Copy number variation: New insights in genome diversity, Published in Advance June 29, 2006, doi: 10.1101/gr.3677206 Genome Res. 2006. 16: 949-961 http://genome.cshlp.org/content/16/8/949.full.html#ref-list-1
10. Hieronymus, H., et al, Copy number alteration burden predicts prostate cancer relapse, PNAS, Vol 111, No 30, July 2014.
11. Iwata, T., MYC Overexpression Induces Prostatic Intraepithelial Neoplasia and Loss of Nkx3.1 in Mouse Luminal Epithelial Cells, PLOS Feb 2010, Vol 5 Issue 2.
12. Kindu, J., Y Surh, Inflammation: Gearing the journey to cancer, Mutation Research 659 (2008) 15–30.
13. Kocerha, R., REGULATION OF THE STEROIDOGENIC ACUTE REGULATORY PROTEIN (StAR) BY cAMP AND TRANSFORMING GROWTH FACTOR-BETA (TGF-BETA) DEPENDENT PATHWAY, PhD Thesis, Univ Florida, 2005.
14. Kassen A, Berges R, Senge T. 2000. Effect of beta-sitosterol on transforming growth factor-beta-1 expression and translocation protein kinase c alpha in human prostate stromal cells in vitro. Eur Urol. 735-741.

15. Lalonde, E., et al, Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study, Lancet Oncology, Nov 2014.
16. Lupski, J., P. Stankiewicz, Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes, PLoS Genetics | www.plosgenetics.org 0627 December 2005 | Volume 1 | Issue 6 | e49.
17. McCarroll, S., Copy number variation and human genome maps, Nature Genetics, 42, 365–366, (2010)
18. McCarroll, S., D. Altshuler, Copy-number variation and association studies of human disease, NATURE GENETICS SUPPLEMENT | VOLUME 39 | JULY 2007.
19. McMenamin, M., et al, Loss of PTEN Expression in Paraffin-embedded Primary Prostate Cancer Correlates with High Gleason Score and Advanced Stage, CANCER RESEARCH 59, 4291–4296, September 1, 1999
20. Mikropoulos, C., et al Translating Genetic Risk Factors for Prostate Cancer to the Clinic, http://www.medscape.com/viewarticle/830689, Future Oncol. 2014;10(9):1679-1694
21. Miller, D., et al, Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies, The American Journal of Human Genetics 86, 749–764, May 14, 2010 749.
22. Pinkel, D., D. Albertson, Array comparative hybridization and its applications in cancer, Nature Genetics, Vol 37, June 2005.
23. Pinto, D., et al, Copy-number variation in control population cohorts, Human Molecular Genetics, 2007, Vol. 16, Review Issue 2
24. Redon, R., et al, Global variation in copy number in the human genome, Nature, Vol 444|23 November 2006| doi:10.1038/nature05329
25. Vergara, I., et al, Genomic “dark matter” in prostate cancer: exploring the clinical utility of ncRNA as biomarkers, Front. Genet., 22 February 2012 | doi: 10.3389/fgene.2012.00023.
26. Waddell, M., et al, Predicting Cancer Susceptibility from Single Nucleotide Polymorphism Data: A Case Study in Multiple Myeloma, Proceeding BIOKDD '05 Proceedings of the 5th international workshop on Bioinformatics Pages 21 – 28.
27. Wojno, K., et al, Reduced Rate of Repeated Prostate Biopsies Observed in ConfirmMDx Clinical Utility Field Study, Am Health Drug Benefits. 2014;7(3):129-134
28. Yu, Y., et al, Genome Abnormalities Precede Prostate Cancer and Predict Clinical Relapse, The American Journal of Pathology - June 2012 (Vol. 180, Issue 6, Pages 2240-2248, DOI: 10.1016/j.ajpath.2012.03.008). http://www.journals.elsevierhealth.com/periodicals/ajpa/article/S0002-9440%2812%2900241-6/abstract

10 RELATED WHITE PAPERS

The following are related White Papers (available on www.telmarc.com) which we have written in the recent past which may facilitate the material contained herein.

No. 119 SNPs and Prostate Cancer

No. 118 Vitamin D and Prostate Cancer

No. 117 SPDEF, ETS Transcription Factors and PCa

No. 116 Methylation, Prostate Cancer, Prognostics

No. 115 Endosomes and Melanoma

No. 114 NOTCH, miR-146a and Melanoma

No. 112 Prostate Cancer: miR-34, p53, MET and Methylation

No. 111 CRISPR and Cancer

No. 110 ERG and Prostate Cancer

No. 108 Cancer Cell Dynamics

No. 107 Prostate Cancer Genetic Metrics

No. 106 Divergent Transcription

No. 104 Prostate Cancer and Blood Borne Markers

No. 103 Prostate Cancer Indolence

No. 102 MDS and Methylation

No. 101 Exosomes and Cancer

No. 100 lncRNA and Prostate Cancer

No. 99 SNPs and Cancer Prognostics

No. 98 CCP and Prostate Cancer

No. 97 ATF2 and Melanoma

No. 96 PD-1 and Melanoma Therapeutics

No. 95 MER Tyrosine Kinase Receptors and Inhibition

No. 94 Melanoma Therapeutics

No. 93 Cancer Cell Dynamics Methylation and Cancer

No. 91 Methylation and Cancer

No. 90 Telomeres and Melanoma

No. 89 miRNA and Melanoma

No. 88 Extracellular Matrix vs. Intracellular Pathways

No. 87 Prostate Cancer Prognostic Markers

No. 86 Cancer Models for Understanding, Prediction, and Control

No. 85 Prostate Cancer Stem Cells

No. 84 Epistemology of Cancer Genomics

No. 83 Prostatic Intraepithelial Neoplasia

No 82 Prostate Cancer: Metastatic Pathway Identification

No 81 Backscatter Radiation and Cancer

No 80 PSA Evaluation Methodologies

No 79 The PSA Controversy

No 77 Obesity and Type 2 Diabetes: Cause and Effect

No. 61 Type 2 Diabetes: A Controllable Epidemic (March 2009)