Human Solid Cancer Decoded

Lan Ko

From the Georgia Cancer Center, Department of Pathology, Medical College of Georgia, Augusta University, Augusta, Georgia, USA

Running title: *GT198 as a cause of human common solid tumors*

Address correspondence to: Lan Ko, MD/PhD, Augusta, GA, USA, Email: lan-ko@outlook.com

Keywords: Unified theory of cancer, pericyte cancer stem cells, cancer diagnosis and treatment, reconciling controversies in cancer research

Abstract:

The secret of human solid tumors was largely concealed by a chicken-egg paradox. Common solid tumors are initiated in tumor microenvironment, and from rare mutant stem cells. A cancer-initiating gene is the only specific marker to reveal mutant cell lineages in tumor stroma, and these cell lineages are required for analysis to reveal the cancer-initiating gene. For decades, it is hard for scientists to find one without first finding the other. Two major types of mutant cell lineages are now revealed. They are mutant immune cells when tumors have chronic inflammation, and mutant blood vessel pericytes when tumors have angiogenesis, often both concurrently present in solid tumors. A DNA repair factor GT198 (gene symbol *PSMC3IP*) is a specific marker of mutant pericyte stem cells and their decedent lineages in solid tumors. Supporting to this concept, here we find that GT198 is a direct protein target of existing anticancer chemotherapy drugs including doxorubicin, paclitaxel, and etoposide. Importantly, GT198 is also a high affinity target of two herbal medicines known to be clinical effective. Therefore, common solid tumors could be systematically treated with low toxic and high efficacy natural herbs. A unified theory of cancer reconciling historical controversies in cancer research is discussed. New methods in cancer early diagnosis and drug development are outlined.

Section I. Introduction

A central challenge in human cancer research in the past is that the true cancer-initiating factor in human common solid tumors was unknown. The lineage identity of cancer-inducing cells was largely unclear. While cellular functions such as DNA repair, cell cycle, apoptosis, and angiogenesis are all eligible consequences of cancer, their true cause was hidden. This mishap prevented the advance of clinical translation, so that early cancer diagnosis is difficult in the absence of cancer initiation marker, and treatment methods remain less effective in the absence of causative target.

Human cancers can be largely grouped into three categories. A category of common solid tumor includes the cancers in breast, ovary, uterus, fallopian tube, prostate, bladder, stomach, head and neck, testis, lung, brain, skin, kidney, thyroid, and colon. These cancers are initiated from mutant stem cells in tumor microenvironment, which is the tissue stroma surrounding the tumors. Tumor cells themselves are only the consequence, being induced, rather than the cause of cancer, so that analyzing tumor itself has missed the identification of cancer-initiating genes. In this article, we will primarily focus on solid tumors.

The second category is called liquid cancer or blood cancer such as leukemia, lymphoma and myeloma. Because there is not a feasible pathology method to distinguish mutant cancer stem cells from normal stem cells and non-mutant tumor cells when they mix together, it lacks a marker to analyze mutant cancer stem cells before knowing the cancer-inducing gene itself. We speculate that the

cancer-inducing gene in blood cancer is a hematopoietic stem cell regulator.

The third category includes sarcomas and rare childhood tumors. In this case, mutant cancer stem cells themselves mature into mutant tumor cells so that analyzing tumor cells have revealed causative cancer genes in the past. At early stage of cancer research, this group of cancer genes were mostly confirmed by cancer genetic studies, such as Rb, EWS, and SYT. However, they are not involved in human common solid tumors.

In solid tumors, it turned out that cancer genes and affected malignant stem cells are hidden in a chicken-egg paradox, only realized after they are decoded. The genes and the cells are like chicken-egg, a cancer gene is the marker to isolate mutant stem cells, and the mutant cells are needed for analysis to reveal the cancer gene. Decades ago, our group have cloned two transcriptional coactivators, RBM14 and GT198, during the study of gene activation. We were essentially the only group focusing on the cancer aspects of these two coactivators, which we now consider them as the two major cancer-inducing genes and the specific markers for cancer stem cells in solid tumor. By originally cloning the cancer genes before knowing them, our group had a chance to first bypass this paradox and unlocked the secret of common solid cancers.

The RBM14 gene at chromosome 11q13 is gene amplified in immune cells in all inflammatory tumors, but its protein product expression is not tumor-specific. RBM14 may be a diagnostic marker but may not be a suitable drug target.

The GT198 gene at chromosome 17q21 is mutated in pericyte stem cells in tumors with angiogenesis. Its protein expression is highly tumor-specific. Essentially pericyte-derived cell lineages affect all solid tumors. This creates a unified cancer concept in which common solid tumors can be systematically treated using inhibitors targeting to mutant cells expressing GT198 protein.

In mouse embryoid bodies, RBM14 is at the first step and GT198 at the second step of stem cell differentiation. Their stem cell regulations may explain their selections as top oncogenes.

In this article, we provide preliminary but hard evidence to show that GT198 is actually a direct target for many FDA-approved oncology drugs including doxorubicin analogs and etoposide. Importantly, clinical effective anticancer herbs can now be confirmed containing inhibitors of GT198. This is an unprecedented opportunity to identify natural herbal medicines with low toxicity and high efficacy for the treatment of human common cancer. Immunotherapy targeting to GT198 and organic purification or screening chemical compound inhibitors are all feasible for the future drug development.

Below we first discuss the hidden cause of common solid tumor, then describe a unified theory in cancer, and further outline the new methods for cancer early diagnosis and drug development. To achieve a cancer breakthrough in the near future, the research, industry and healthcare communities shall be united. Patient survival is the final proof.

Section II. Define the Cause of Human Common Solid Cancer

The evidence to define a first-hit oncogene encompasses cancer genetics, biochemistry, tumor microenvironment pathology, drug development, and immunotherapy. Each section is discussed below:

1) Cancer Genetics

The original notion in quest of cancer gene to end cancer still stands today. The genes that initiate cancer have been found (1,7,8), but the question remains how we prove them. The previous two gold standards in cancer genetics are recurrent somatic mutations and germline mutations with segregation (9-12). These standards indeed had resulted in many oncogenes in sarcomas or rare childhood tumors, but not in common solid tumors (13-15). The common cancer genes were mysteries, such as BRCA1 somatic mutations were hardly present in tumors (16), and p53 mutations were not present at initial stages. This mishap prevented the advance of tumor biology into translation. While DNA repair, cell cycle, apoptosis, angiogenesis, and many modern focuses are all eligible consequences of cancer, their true cause was hidden. Hence, early diagnosis is difficult, and cancer treatment remains less effective, without a causative marker or target.

We now understood the difference between rare and common cancers. In childhood tumors, fast growing tumor cells are direct descendants of mutant stem cells, so that they possess recurrent

somatic mutations such as EWS, SYT, WT1 or Rb. In common solid tumors, mutant stem cells are immune cells or blood vessel pericyte stem cells, so that epithelial tumor cells do not possess mutations in the first-hit cancer genes, whereas only tumor stroma does (see section 3 below). Therefore, this was a chicken-egg paradox. Without knowing cancer genes, there was no marker to find mutant stromal cells, hence no avenue to sequence common cancer genes out of rare mutant stromal cells using genetic approach.

The next option left was germline mutations with segregation to prove a common cancer gene as first-hit. Leading cancer geneticists have previously identified major cancer loci including chromosome 11q13 (17-19), and 17q21 (20-22), where the two hidden cancer genes actually locate (Figure 1). In the 11q13 locus, while MEN1 losses copy numbers and CCND1 is amplified, in between a gene cluster of RBM14 (also called CoAA) (5), losses its enhancer with amplified and unregulated gene body (Figure 1A)(1). RBM14 might have few cancer-associated families, since it was quite unlikely to permit embryonic development into an adult when carrying hundreds copy of amplified genes. In the 17q21 locus, the presence of a very small GT198 gene was shadowed by the finding of BRCA1 just 470 Kb away, and was missed during early genetic identification (Figure 1B) (7,16,23-25). Both RBM14 and GT198 are located within fragile genomic regions between well-studied cancer genes with gain and loss (Figure 1). GT198 is also amplified in breast cancer (7).

It appears that early genetic studies keenly focusing on copy gain or loss regions have overlooked the genes in between, where genetic stability could be severely compromised. In addition, another unrecognized pitfall is that germline and somatic mutation frequencies are often reciprocal. Cancer genes control stem cell differentiation and embryonic development. A deleterious mutation would leave fewer individuals to survive in pedigrees but is prone to be prevalent in tumor due to higher functional impact. BRCA1 and BRCA2 have larger pedigrees and rare somatic mutations (26). GT198 or p53 have smaller pedigrees and abundant recurrent somatic mutations (2,4,6-8). Gene amplification of RBM14 in common solid tumors can reach 90-100% (Figure 3A), and no family may carry it. In contrast, disease genes and SNPs have larger associated families due to less impact on embryonic development. Thus, analyzing large cancer pedigrees for experimental confidence was a grave pitfall, and resulted in historical failures in finding hidden first-hit cancer genes through cancer genetics approach.

2) Biochemical Philosophy in Cancer Stem Cells

This chicken-egg paradox in cancer gene was finally unlocked through biochemistry rather than cancer genetics. Similar to cancer genetics, biochemical logic can stand alone in the absence of the input from tumor biology. The basic principle in biochemistry is that "structure defines function."

In a transcriptional study aiming to find top regulators surrounding RNA polymerase II (Pol II) (Figure 2A), a protein factor initially called CoAA (gene symbol RBM14) was identified (27). RBM14 is an RNA-binding protein containing highly repetitive tyrosine-rich sequences, termed YxxQ repeats (Figure 2B-C). Surprisingly, in the entire

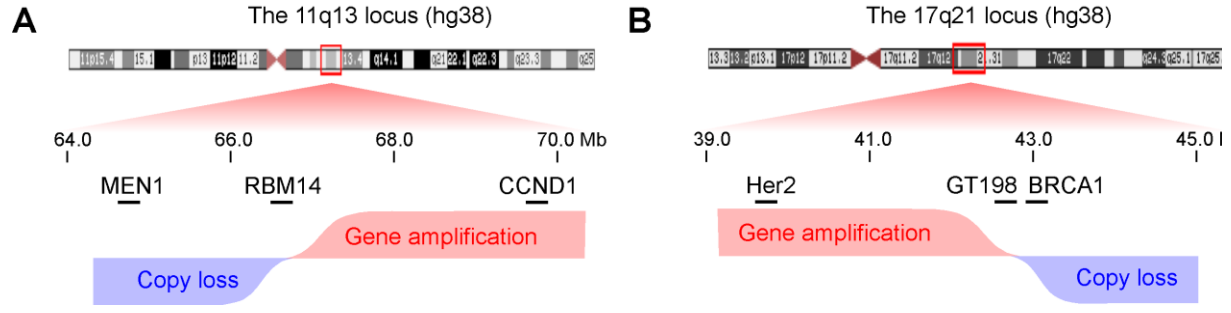

Figure 1. Chromosome 11q13 and 17q21 loci. (A) The RBM14 gene lost its enhancer while multiplied its unregulated gene body during gene amplification at 11q13 (1). **(B)** The GT198 gene is located at 470 kb proximal to BRCA1 at 17q21. The Her2 (ERBB2) gene is 2.9 Mb away from GT198.

human genome analyzed by protein motif alignment, only a few proteins have the same type of YxxQ repeats. All of which are established oncoproteins (EWS, TLS, TAF68, SYT, BAF250) (1,28). This immediately put RBM14, at 11q13, as an oncogene candidate based on the notion of structure defining function. Gene amplification of RBM14 was subsequently found to be prevalent in common tumors (Figure 3A) (1). Recurrent gap deletions revealed by qPCR were found in three unrelated tumor specimens (unpublished).

In addition to genetic evidence, the second most convincing evidence is that RBM14 is a direct Pol II-interacting protein, the only one of this kind in the human genome. The Pol II C-terminal domain (CTD) has similar tyrosine-rich diheptad repetitive motifs required for a zipper-like interaction (Figure 2B-C) (29-31). Mutagenesis at both Pol II and RBM14 tyrosine-rich motifs confirmed phosphorylation-dependent interaction *in vitro* (32). EWS and SYT are known Pol II-interacting proteins (33). BRCA1 also interacts with Pol II directly (34). The entire DNA repair complex is part of the transcriptional complex thereby interacting with the Pol II. Thus, the top secret of cancer genes lies in the Pol II cabinet (Figure 2A). They represent a control center to govern a pyramidal organization below them. This logical arrangement ensures a single RNA Pol II enzyme capable to transcribe the entire human genome on a specific gene, in a specific cell, and at a specific time. Only faulty actions at the top will result in destructions beyond repair in cancer (15).

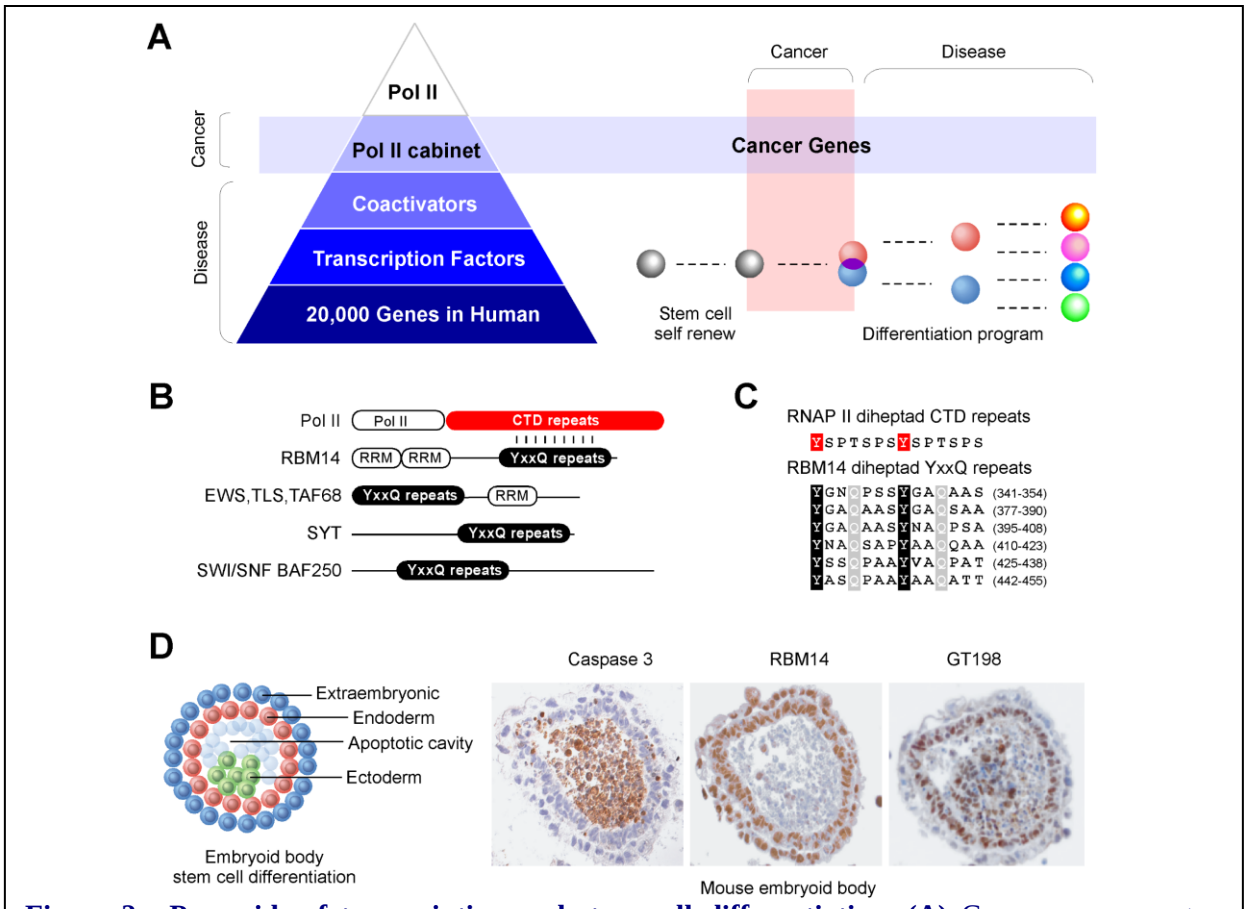

Figure 2. Pyramids of transcription and stem cell differentiation. (A) Cancer genes are top transcriptional regulators encoding Pol II-interacting proteins. They control the initial or early steps of stem cell differentiation. **(B)** Oncoproteins identified through ScanProsite sequence alignment of YxxQ motifs (1), which interact with the CTD of Pol II. **(C)** Diheptad Pol II CTD repeats and RBM14 YxxQ repeats. **(D)** Immunohistochemical staining of RBM14, GT198, and Caspase 3 for apoptosis in mouse embryoid bodies. The model at the left indicates the cell layers presence of wild types and variants of oncoproteins (4-6).

The fault at the lower levels will mostly affect pathways in diseases (Figure 2A).

The third critical evidence is that both RBM14 and GT198 regulate stem cell differentiation at initial stages (Figure 2D), and through their spliced variants (4,35). The RBM14 orthologue, Lark, in *Drosophila* controls circadian rhythm (5,36). Cancer is a stem cell disease in which mutant stem cells are unable to terminal differentiate (37). Otherwise, tumors would be replaced during normal homeostasis. It has been demonstrated that cascade alternative splicing is the key in step-wised stem cell differentiation (38-41). A wild type protein enters one daughter cell while its functional opposing splicing variant enters the other, so that asymmetric cell division produces distinct offspring (Figure 2A) (42).

A fundamental step of cell differentiation, in normal, cancer, or disease, is a segregation of Ying-Yang. Ying and Yang are defined as opposite activities or forces that are mutually dependent, mutually inclusive, coexisting, and exchangeable, such as the concepts of hot and cold. In cells, regulatory genes produce splice variants as counterparts. The Yang wild type and Ying variant segregate during cell differentiation and produce distinct offspring. The failure of this segregation in stem cell causes cancer, and in differentiated cells causes disease. Hence, normal, cancer, disease are unified under the same principle. This Ying-Yang segregation is a fundamental philosophy to stem cell differentiation. Most known Pol II cabinet proteins possess the same characteristics having counter splice variants (RBM14, GT198, p53, BRCA1, EWS) (4,5,43-45). Mutations cause their variants increase or over activated in cancer.

In an embryoid body paralleling embryonic development, RBM14 first segregates with its apoptotic splice variant into two layers of cells, so that it may cause a center cavity to form (Figure 2D) (5). Then, GT198 expression distinguishes the endoderm layer (Figure 2D) (4). Multiple subsequent factors continue in the same fashion at later steps of differentiation. It is conceivable that Pol II cabinet proteins are dominant at early steps so that a chain reaction of target genes will follow. This theory implies a critical role of Pol II cabinet in adult tissue stem cells, such as hematopoietic or vessel pericyte stem cells. In fact, alternative splicing switch in stem cells was the most decisive reason prompted us to re-focus on GT198 in cancer in 2009, despite its cloning in 2002 (25), foreseeing overexpressed GT198 versus amplified RBM14 as a better cancer drug target.

The fourth critical evidence, in addition to the evidence from our group, is that other groups have extensively shown that GT198, also called Hop2 or TBPIP, regulates recombination, meiosis, and DNA repair (46-50). It was clear that GT198 is a critical factor in the DNA repair complex which serves as Pol II cabinet. GT198 is a DNA-binding protein dimer similar to p53. GT198 is capable to interact with both single and double-stranded DNA in a non-sequence specific manner (4), consistent with its action in transcription, DNA recombination and repair. An ideal strategy in drug development is to target GT198 DNA-binding activity, which in fact, has already inhibited by existing FDA-approved clinical oncology drugs (see section 4 below).

3) Pathology of Tumor Microenvironment

Adult tissue stem cells can be hematopoietic or perivascular and are located in connective tissues known as stroma (in contrast to parenchyma). Hematopoietic stem cells produce immune cells. Vascular pericyte stem cells produce stromal vascular smooth muscle cell lineage, such as fibroblasts and adipocytes. Tissue stroma is also a natural home of adult stem cells. Normal stem cells, when mutated, become cancer stem cells. They produce descendent cell lineage that is capable to drive epithelial tumor cell growth. For example, in ovarian cancer, surrounded by far more number of normal stem cells, mutant stem cells differentiate into mutant hormone-producing theca cell lineage to stimulate epithelial cell growth (6). Thus, common cancer initiates in stroma before tumor cell appears. That is why tumor microenvironment host early changes in cancer.

This was a deep problem for pathologists who traditionally only look at tumor cells. This was also a hurdle to geneticists who analyze tumor mass rather than tumor stroma. This caused frustration to cell biologists who analyze mostly tumor-derived cell lines. This confused immunologists who detected counter immune responses without knowing that both healthy and mutant immune cells coexist, and smoking and virus infection are just risk factors for creating inflammation-induced mutant immune cells. This burdened stem cell biologists whose stromal theories met resistance

from tumor biologists. This restricted angiogenesis experts who consider blood vessels normal rather than malignant (3). This ultimately hindered clinical oncologists since drugs targeting tumor lead to the escape of mutant stromal cells until metastasis. Finally, this tumor stroma biology prevented the scientific community to recognize the two common cancer genes in the past decades.

Gene amplification of RBM14 is prevalent in tumor stroma (Figure 3A). In particular, tumors associated immune cells have the highest frequency. Tumor associated lymph nodes are almost affected 100% by RBM14-amplified immune cells, implicating metastatic spread of mutant immune cells. Often, the copy number in each cell can reach hundreds (Figure 3A). Some amplified immune cells carry markers such as c-kit CD34 and CD133 (not shown), although their precise lineage identity has not been resolved.

The tumor-associated GT198 overexpression is cytoplasmic, in contrast to its normal nuclear expression in embryonic tissues (2,3,6). Affected

A

RBM14 Gene Amplification in Cancers

Cancer type	*Positive/Total*	*%*
Lung	44/59	73
Skin	45/50	90
Lymphoma	60/78	77
Hodgkin's	4/4	100
Spleen	13/13	100
Thymus	4/4	100
Head&Neck	29/30	97
Stomach	6/10	60
Esophagus	9/10	90
Colon	9/10	90
Thyroid	5/10	50
Bladder	7/10	70
Cervix	2/10	20
Pancreas	7/10	70
Brain	5/6	83
Placenta	0/8	0

B

GT198 Stromal Expression in Cancers

Cancer type	*Positive/Total*	*%*
Breast	26/60	43
Ovarian	13/60	21
Uterine	10/50	20
Fallopian	19/30	63
Prostate	43/62	69
Bladder	13/20	65
Testis	3/10	30
Lung	16/28	57
Brain	15/18	83
Melanoma	6/18	33
Kidney	3/11	27
Oral	33/40	82
Thyroid	4/10	40
Colon	8/18	44

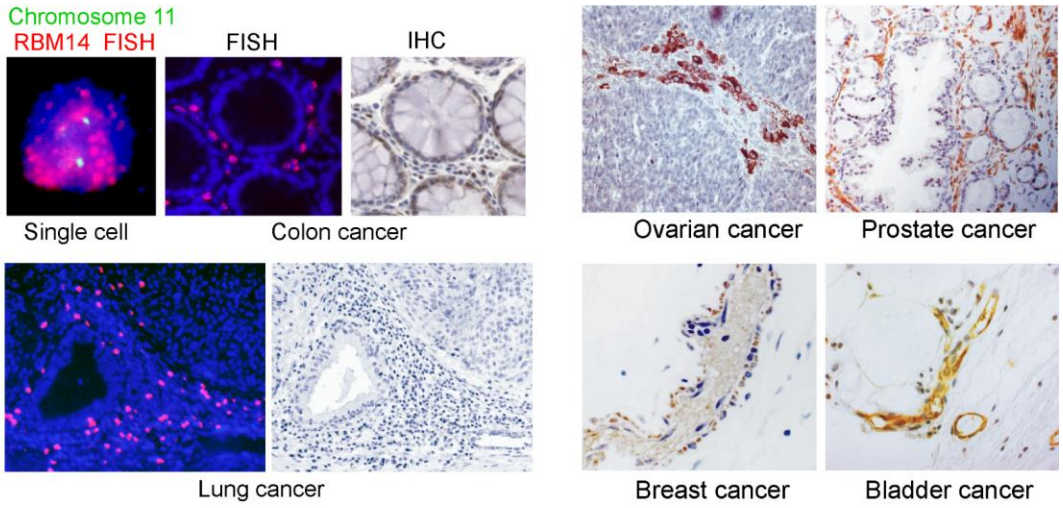

Figure 3. RBM14 gene amplification and GT198 cytoplasmic overexpression in human tumor stroma. (A) FISH analysis of RBM14 gene amplification in tumor arrays. Representative images of colon and lung cancers showing RBM14-positive stromal cells. **(B)** Immunohistochemical staining of GT198 in tumor stroma. Representative images of ovarian, prostate, breast, and bladder cancers showing GT198-positive stromal cells (2,3). Positive rates are potentially higher if not analyzed in small-sized tumor arrays but in large-sized tumor slides.

cells include pericytes and vascular smooth muscle cell lineage. These include myoepithelial cells and adipocytes in breast cancer (2), theca cells in ovary cancer (6), myofibroblasts in prostate and bladder cancers (Figure 3B), and pericytes in multiple other common cancers (3). GT198 somatic mutations are abundant in positive breast and ovarian cancer stroma (2,6). High frequency somatic mutations are also present in prostate and bladder cancer stroma (not shown).

The blood vessels in tumor angiogenesis were previously considered normal. Our evidence indicates that vessel pericytes in angiogenesis are malignant, overexpressing GT198 (3). These pericytes evolve into tumor cells resembling the local tissue types such as squamous cells in oral tumor, or glioma cells in brain cancer (3). These pericyte-derived tumor cells also migrate into tumor-associated lymph nodes suggesting "cancer stem cells" responsible for tumor metastasis (3). Thus, a metastatic tumor does not necessarily resemble the original tumor as often observed, but is more compatible to distant home environment. The presence of GT198-induced malignant vessels confirms the long-standing notion of tumor angiogenesis as a cause of common cancer. Since the cancer gene GT198 is the cause of tumor angiogenesis.

4) "Nature as A Remarkable Chemist" and A Doctor

Patient survival is the only proof. If GT198 serves as a cause of cancer, drugs targeting to GT198 may be effective. To test this idea, we screened NCI FDA-approved chemotherapy drug collections, using *in vitro* plate binding assay to detect direct inhibition of DNA-binding of GT198. When DNA is biotin-labeled, the bound DNA can be measured using streptavidin-horseradish peroxidase (HRP) conjugate under the inhibition of various concentrations of tested drugs. Surprisingly, we identified many positive GT198 inhibitors including all doxorubicin family members, and mitoxantrone (IC50=149-973 nM) (Figure 4A-C). Camptothecin, topotecan, and irinotecan had same but very poor affinity (IC50=2.6 μM). Etoposide had much higher affinity at IC50=24.2 nM (Figure 4B). However, this is no longer surprising when GT198 was found to share protein sequence homology with both DNA topoisomerases I and II (Figure 4D), which are previously considered as targets for these drugs. Platin DNA inhibitors did not inhibit GT198 directly (Figure 4B). Yet their combined effects on the DNA-binding protein GT198 remain to be tested. The positive ones are in fact among clinical effective cancer drugs. These results immediately confirm that GT198 is a drug target.

We also tested GT198 with paclitaxel, which is a natural compound originally isolated from the Pacific yew tree (51,52), and a successful clinical drug. The mechanism of action on microtubules in mitotic arrest, which is ubiquitously present in all cells, does not explain the clinical efficacy or side effects of paclitaxel (53). The original publication in 1981 showed paclitaxel binding to microtubules at 870 nM (54), hundreds fold weaker than the binding affinity to GT198 at 5 nM (Figure 4B). However, paclitaxel may have low efficacy due to allosteric inhibition of GT198, in contrast, doxorubicin competitively inhibited GT198 (tested by binding kinetics in double reciprocal plots and data not shown) with higher efficacy but lower affinity (Figure 4B).

From a clinical perspective, our new finding provides a rationale for the clinical side effects of paclitaxel. In human cancer patients, paclitaxel-induced lymphocyte toxicity, neutropenia, testicular shrinkage, and ovarian damage, are potentially correlated with the endogenous GT198 protein expression specifically in testis, ovary, spleen, and thymus (3,25). The acute paclitaxel toxicity in rodents was also found in the ovary (55), testis, hematopoietic, and lymphoid systems (56). In contrast, the low toxicity of paclitaxel in major organs is correlated with undetectable GT198 protein level in most of these tissues (3,25). The GT198 protein is also expressed in neurons which may explain paclitaxel neurotoxicity (3). Furthermore, inhibition of DNA repair protein activities by paclitaxel has been shown by others (57). Together these mechanistic and clinical observations strongly support that GT198 is potentially one of the direct targets of paclitaxel.

It is also important to note, that GT198 overexpression is cancer-specific (3), in contrast to ubiquitous DNA topoisomerases or microtubules. GT198 is therefore a previously unanticipated drug target. It is possible that clinical drugs selected from cells and animal assays were inhibitors of multiple cellular proteins including but not limited to GT198, which explains that highly cytotoxic

camptothecin was not clinical successful. The correlation between clinical drug efficacy and GT198 expression in cancer needs to be more detail analyzed to assess drug efficacy and toxicity. Importantly, with GT198 emerged as a direct drug target, it is now possible to identify GT198 inhibitors with both high affinity and efficacy.

Inspired by the discovery of Artemisinin (58), as well as paclitaxel (52), we screened anti-tumor herbal medicinal plants and found herbal drugs *Gleditsia Sinensis Lam*. and *Pimenta Dioica* containing GT198 inhibitors (Figure 4E).

Gleditsia Sinensis Lam. (皂角刺 in Chinese) is one of the alleged 50 fundamental herbs used in traditional medicine, and has been used in thousand years with antitumor effect. The thorn needle extracts from this plant (Figure 4E), were described to have effects specific to hormone-associated tumors such as breast, ovarian, and prostate cancers, which link to GT198 functions. In addition to ancient practices, currently *Gleditsia Sinensis* is widely used as anti-breast cancer herb in Asian clinics, and the tree is massively planted. Modern research on *Gleditsia Sinensis* is increasingly significant and also suggests that it has effects on breast and prostate cancer (59,60), cell proliferation and cell cycle (61), and angiogenesis (62-64). Its biological activity is recently summarized (65).

The second plant we identified is called *Pimenta Dioica* (also known as Allspice or Jamaican pepper). The aromatic berries of allspice are a common spice originally introduced to Europe from Jamaica by Columbus. Folklore describes it as anti-cancer, and allspice is considered a highly organic food. Importantly, a long time use does not have any observed adverse effects if the dose consumed as a spice. The allspice extracts have previously been shown as a powerful anti-prostate

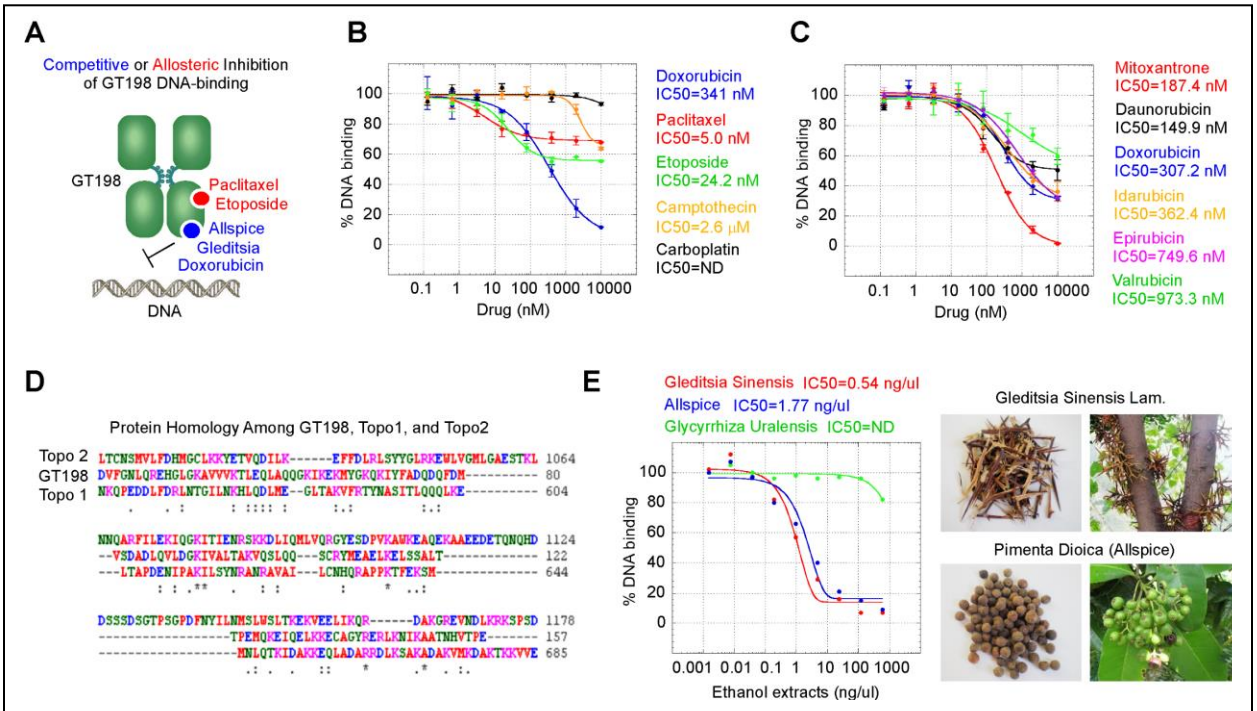

Figure 4. GT198 is a drug target. (A) Hypothetical model of competitive (blue) or allosteric (red) inhibition of GT198 dimer in blocking the DNA-binding. **(B-C)** Inhibition of biotin-labeled DNA binding to GT198 under increasing concentrations of clinical drugs as indicated. IC50s are measured by sigmoidal curve fit. **(D)** Primary sequence homology among GT198 and DNA topoisomerase I (Topo 1) and II (Topo 2) (Clustal Omega). **(E)** Methanol extracts from thorn needles of Gleditsia Sinensis and berries of Allspice (picture on the right) inhibit GT198 DNA-binding activity at high affinity (IC50=0.54 and 1.77 ng/μl) and high efficacy (90% inhibition). A common herbal medicine Glycyrrhiza uralensis (Chinese liquorice root) was used as a negative control.

cancer drug (66). It appears that multiple active ingredients may exist in the allspice extracts, which have pro-apoptotic, pro-autophagy, and anti-tumor activity in mouse models (66). Both *Gleditsia Sinensis* and allspice materials are abundant, easy to obtain, and commercially available. This eliminates the environmental concern if chemical compounds need to be isolated from the plants to serve as synthetic substrates for clinical use.

Both *Gleditsia Sinensis* and allspice ethanol extracts inhibit GT198 potently (0.5 and 1.7ng/ul, Figure 4E). The inhibition is competitive rather than allosteric which was confirmed using binding kinetics. The chemical properties of *Gleditsia Sinensis* and allspice are distinct in polarity. If purified, they could serve in combination together to achieve even higher drug efficacy. Thus, purification of these natural products holds high promise to obtain low toxic and high efficacy drugs.

Herbs have been the principal form of medicine since the ancient time, and this is still true in most developing countries. Ancient civilizations in Egypt, Europe, the Middle East, India, China, and Africa have accumulated mounting sophisticated medical uses of herbs, many of which were described in ancient medicinal books. Traditional Indian and Chinese herbal medical books list at least several thousands of plants. It is estimated that around 70,000 plant species have medical use one time or another during the history (67). Many well-known chemical drugs such as aspirin, etoposide, and paclitaxel are all originally identified from plants.

In the history, the example of marketing paclitaxel, originally named as taxol, was unique that could not be repeated today (51,52,68). The project was first initiated by a non-peer reviewed NCI decision, then followed by determined Wani and Wall at RTI, further triggered by mechanistic studies towards DNA functions including mitosis (69), and topoisomerases, which were quite related to GT198 functions. Thereafter, NCI drug formulation paved the way to clinical trials. After semi-synthetic efforts by numerous researchers, and a controversial Bristol-Myers Squibb deal granted by NCI, paclitaxel was finally made to patients' bed in 30 years. Still, paclitaxel has far weaker efficacy and is less abundant than unidentified compounds in two herbs described above. Today, we could do much faster. The current technology in organic chemistry has been greatly improved than the past, and activity can be monitored by *in vitro* assay rather than cytotoxicity, so that new drugs can be soon purified.

5) Immunotherapy

Since oncoprotein GT198 is overexpressed in cancerous stromal cells, and aberrantly in the cell surface (3), GT198 may be an ideal antigen in immunotherapy. We first tested this idea in several mouse tumor models and found promising preliminary results. These include implanted GL261 glioma and 4T1 breast tumor (3), and spontaneous MMTV-PyMT breast tumor (Figure 5).

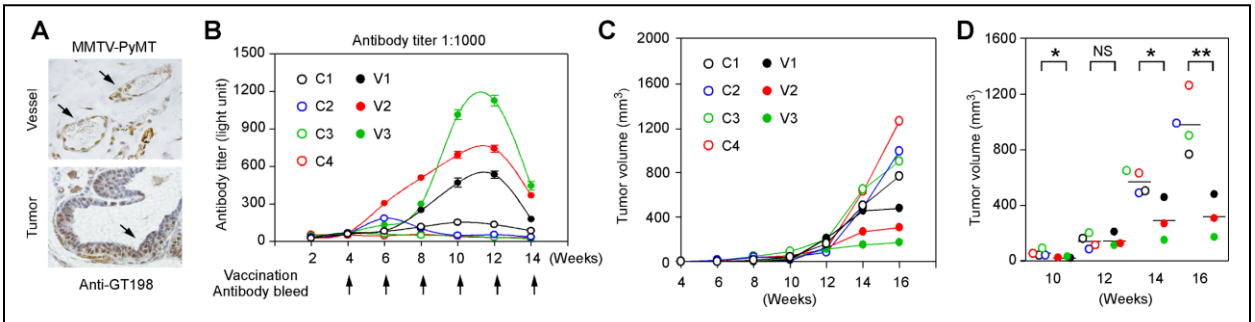

Figure 5. GT198 vaccination in MMTV-PyMT breast tumor model. (A) Immunohistochemical staining of GT198 in MMTV-PyMT spontaneous breast tumor showing positive angiogenic vessels and tumor cells. Arrows indicate GT198^{+} cells. **(B)** MMTV-PyMT mice were subcutaneously vaccinated by injection of recombinant GT198 protein as antigen (100μg) starting at age of 4 weeks. Tail blood serum (1:1000) from each bleed was analyzed for anti-GT198 antibody titer. **(C-D)** Tumor volumes were measured and analyzed in control group (n=4) and vaccine group (n=3).

In PyMT tumors, GT198 is overexpressed in vessel pericytes and breast tumor cells (Figure 5A). Vaccination using recombinant GT198 protein as antigen showed a correlation between higher anti-GT198 antibody titers (Figure 5B), and smaller tumor sizes (Figure 5C-D). Lung metastasis was also suppressed in the vaccine group.

GT198 protein fulfills most criteria of an ideal antigen for cancer vaccine: First, distinct from tumor antigens only in tumor cells, GT198 is mostly expressed in tumor-inducing stromal cells, particularly in pericyte stem cells and progenitors. Eliminating defective stem/progenitors will be expected to remove the stimuli to fuel the tumor growth. Secondly, endogenous GT198 protein is a cancer testis antigen, and is mainly expressed in embryonic tissues and in testis/ovary in adults providing high specificity to minimize toxicity in normal tissues. Thirdly, wild type GT198 is a nuclear protein but overexpressed GT198 in cancer is in cytoplasm and on cell surface (3), adding its eligibility and specificity in immunotherapy. Fourthly, GT198 is a small protein at 217 amino acids enriched with alpha-helices. GT198 has very high antibody titer when polyclonal antibody was produced, indicating a likely success of GT198 antigen with high immunogenicity. Finally, GT198-expressing pericytes are common in angiogenic tumor stroma of human solid cancers (3), so that a successful immunotherapy approach may have high impact in common solid cancer treatment. Thus, targeting GT198 as an antigen will be a promising strategy in cancer immunotherapy.

Section III. Unified Theory of Cancer

1) Unity in Cancer Research

With the emergence of RBM14 and GT198, we now can propose a new gold standard to define first-hit common cancer gene if not previously sufficient. 1) Possess germline or recurrent somatic mutations; 2) function as a Pol II cabinet protein or in the DNA repair complex; 3) regulate stem cells using alternative splice variant; 4) mutated or altered in tumor stromal stem cells; 5) clinical effective if targeted by its inhibitors or antibodies.

The above standard encompasses multiple disciplines including cancer genetics, biochemistry, pathology, stem cell biology, immunology, and clinical sciences. It should not be too surprised that cancer research has not been unified. Most of the cancer research scientists are primarily trained in one or two narrowed fields, and are increasingly more focused when they become seasoned experts. The interdisciplinary study thus becomes harder to be conducted due to the reviewing system known to be problematic (70). In cancer research, when scientists are restricted only to reviewer-approved existing directions and must propose conclusion before experimentation, innovation is restricted.

Cancer research is an analogy of the "six blind men and the elephant" (Figure 6). In the absence of a unified concept, decades of cancer research has generated diversified ideas sparking inconsistency and dispute. Since any single evidence is part of the integrated whole, multidisciplinary studies are required to reconcile the evidence. Yet this is uniquely challenging when traditionally reviewing system often focuses on single research discipline. This challenge further amplifies for the fact that drug development is a collective effort encompassing academia, industry, clinical science, and government sectors. To reconcile, cancer study must be broadened as well as simplified. Albert Einstein has once said: "Out of clutter, find simplicity. From discord, find harmony. In the middle of difficulty lies opportunity."

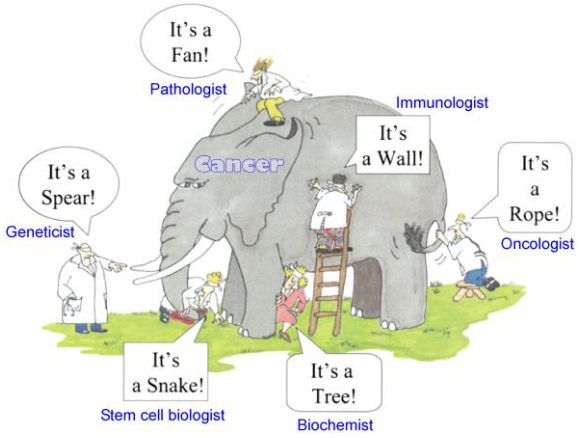

Figure 6. Six blind men and the elephant. Scientists are capable of discovery in a single field, but each view is inherently limited by the failure to account other truths as an integrated whole. To uncover the secret of cancer, broadened and simplified vision is needed to reconcile multidisciplinary studies in cancer research. (Cartoon adopted from Marilyn Mehlmann with modifications).

2) Reconcile Orthodoxies in Cancer Research

We now have an opportunity to explain some of the previous controversial observations using the RBM14 and GT198 genes, although far more studies are needed to resolve and reconcile them.

1. Tumor suppressor gene versus oncogene. The original belief was that one copy of the gene can produce 50% activity (actually based on an enzyme) so that a loss of gene copy is a functional decrease. When p53, GT198, and many other tumor suppressors were found "gain" in functions, via splice variants, the two concepts of tumor suppressor gene and oncogene blurred. The presence Ying-Yang concept in gene regulation through splice variants, imply that these two types of cancer genes are actually one.

2. BRCA1 and p53. The mysterious p53 did not affect many tumors at early but mostly in advanced stages. This is possibly due to its ubiquitous rather than stem cell-specific expression pattern, despite being a true cancer gene. In contrast, GT198 and BRCA1 have similar tissue expression patterns in embryos or stem cells permitting their defects in early cancer (3,71). Their similar expression patterns imply the existence of a potentially shared enhancer for BRCA1 and GT198 within the 17q21 locus, which may link to unexplained breast cancer families in historical genetic studies. Many breast cancer families in fact often had relatives with other common cancers, since GT198 impact on common solid tumors.

3. Herceptin. It is well-known that Herceptin clinical efficacy is inconsistent with Her2 protein expression. GT198 and Her2 co-expression was found in breast and ovarian cancer tumor stromal cells (not shown). Given that GT198 and Her2 genes are only 2.9 MB apart at 17q21, GT198 may be inside the Her2 amplicons. GT198 is also known to be amplified in breast cancer (7). Thus, GT198 positive stromal cells could be targeted by Herceptin when Her2 and GT198 co-amplify or co-express. GT198 and Her2 co-expression deserves in-depth study to resolve Herceptin clinical inconsistency.

4. DNA inhibitor drugs. In early drug development, there are many selected drugs are DNA inhibitors, despite of the fact that every cell has the same amount of DNA. The reason that DNA inhibitors can be cancer drugs possibly because the true targets are DNA-binding proteins such as GT198. The DNA inhibitors may sensitize other inhibitors in blocking DNA-binding oncoproteins.

5. Clinical side effects. Many common side effects of chemotherapy drugs can now be partially explained by the endogenous expression patterns of GT198 (3). GT198 is expressed in normal bone marrows, blood cells, certain neural cells, and testis. Thus, it is very critical to apply high standards in selectivity, binding affinity, and inhibition efficacy in screening GT198 inhibitors to offset toxic effects. Natural product inhibitors hold high promise.

6. The double-edged sword of the immune response. It was a central problem in immunology that immune response is needed to against cancer, but meantime needs to be suppressed in cancer-associated chronic inflammation. We now know that RBM14-amplified mutant immune cells coexist with healthy immune cells. If mutant ones can be selectively targeted in cancer therapy, rather than suppressing a marker on a group of normal immune cells in the body, cancer therapy could be more effective. It will be interesting to test whether DP-1 is expressed on mutant RBM14-amplified immune cells.

7. Cultured cell lines and animal models. Starting from HeLa cells, cancer cell lines have been most extensively studied in cancer research, but did not always deliver satisfactory results. We observed that all fast cycling cultured cells, even normal cells, have a high level of GT198 protein (25). In human cancer, however, GT198 expression in tumors can be absent or often in tumor stromal cells (2,6). Thus, tumor-derived cancer cell lines do not reflect real tumors in human. In mouse tumors, GT198 expression varies (3). The Ras oncogene-induced mouse tumors have high GT198 expression since GT198 was activated in the Ras pathway (25). Thus, historical drug development using cells or animal models only had limited luck. Some successful ones accidentally inhibited GT198. Some unsuccessful ones might inhibit cell cycle or cell survival functions. In addition, when mouse and human genes are compared in GT198 and RBM14, the subtle differences among species, in enhancer or promoter, suggest that human but not mouse is vulnerable to cancer. In this regard, animal models can test tumor development, but not uncover tumor initiation. We have not found GT198 mutation or RBM14 amplification in mouse.

8. Cancer stem cells. We now consider that cancer stem cells are mutant stem cells originate from stroma rather than from tumor. They can be hematopoietic stem cells associated with chronic inflammation, or pericyte stem cells in blood vessels (Figure 3). This view, however, does not exclude the possibility for mutant stem cell spread into the tumor. This view also emphasizes that the source of metastasis is likely the mutant stromal stem cells (3).

Although our current study is a tip of the iceberg in resolving cancer as a whole, our findings provide a unique opportunity to connect and integrate the vast existing evidence, and to build a unified theory of cancer. Our findings in turn point to important new directions in human cancer diagnosis or treatment as described below.

Section IV. Diagnosis and Treatment of Common Solid Cancers

We list a number of future directions in developing diagnostic methods and therapeutic drugs. The technology *per se* is not the focus since the modern technology is more than sufficient as long as collaboration with experts as listed. In fact, simplified and streamlined methods have better likelihood to be successful.

1) Cancer Diagnosis Methods

1. A FISH diagnostic kit using RBM14 BAC clone as a probe to detect RBM14 gene amplification in inflammatory cancers. The positive rate is expected to be 70-100% if using pathology slides. Retrospective clinical pathology sample analysis for FDA approval as a diagnostic kit. — Clinical pathologists.

2. Mapping and cloning RBM14 amplicons to detect gap deletions. A sensitive PCR crossing the gap will be useful to detect trace amount of amplified mutant immune cells from blood DNA of cancer patients. This can be a cancer screening method to detect cancer early or to detect cancer relapse. — Cancer geneticists.

3. An immunohistochemistry (IHC) kit using a GT198 antibody to detect early cancer from pathology specimen. Retrospective pathology analysis for FDA approval as a diagnostic kit. — Clinical pathologists.

4. GT198 mutation sequencing in screening breast and ovarian cancer families (like in BRCA1). Positive rate in breast cancer families is 4-5%. In addition, sequencing mutations in tumor DNA in conjunction with the use of IHC kit can confirm the presence of early cancer. — Cancer geneticists.

5. A FISH kit using both GT198 and Her2 DNA probes to predict Herceptin efficacy. Retrospective clinical pathology sample analysis for a diagnostic kit. — Cancer geneticists and pathologists.

6. Detection of GT198 protein expression in tumor microenvironment to predict clinical efficacy of drugs such as paclitaxel and doxorubicin to facilitate the drug treatment in cancer patients. Retrospective clinical analysis. — Oncologists and pathologists.

2) Cancer Drug Development

1. Immunotherapy using GT198 protein vaccine or anti-GT198 antibody. Identification of a GT198 peptide or a monoclonal anti-GT198 antibody with effective epitopes in cells and mouse models. — Immunologists and mouse experts.

2. Natural product purification. *Gleditsia Sinensis Lam*. and allspice can be purified using organic chemistry for structure analysis and synthesis. — Biochemists and organic chemists.

3. Screening chemical compound library. Since GT198 is a true protein target and the screening assay is sensitive (Figure 4), chemical compound libraries can be conveniently screened. In addition, existing drugs such as paclitaxel or doxorubicin can now be better chemically modified. — Organic chemists.

4. Herb medicine approach. To deliver low toxic drugs to cancer patients sooner, partially or unpurified herb medicines or their combinations need FDA approval. Allspice is an organic food. — Clinical oncologists.

Section V. Summary

The two oncogenes RBM14 and GT198 are considered the first-hit cancer genes in human common solid tumors. Although more critical genes may exist and can be identified in the future, they together will not be too many. All first-hit cancer genes shall be, and so far were found to be, in the Pol II cabinet. Pol II will not directly interact

with too many molecules. Few existing cancer genes fulfill the current new criteria: 1) Possess high frequency recurrent somatic mutations, a theoretical one hundred percent; 2) Interact with Pol II directly and regulate at the top level of transcription; 3) Control stem cell differentiation at early steps; 4) Carry mutations in stem cells in tumor microenvironment; 5) Clinical effective if targeted by their inhibitors or antibodies.

Most cancer treatment approaches aiming to reducing the bulk of tumor cells are clinically effective, since the tumor cells are a primary burden to resolve disregard many mutant stem cells left behind after treatment. However, a risk of cancer relapse is present if cancer-inducing factors are not completely eliminated. In practice, the existing treatment methods shall be combined with proposed new methods to achieve a better treatment outcome.

A breakthrough of cancer relies on the collective efforts from academia, industry, clinical healthcare, and government funding bodies. Innovation in both science and administration is required. Together, it is conceivable to have a cancer cure if the cause of cancer can be targeted.

Acknowledgement

This work was supported in part by the Georgia Cancer Coalition Distinguished Cancer Scholar Award. I acknowledge the help and support from numerous international scientific collaborators in the past.

Methods:

Fluorescent in situ hybridization (FISH)

Institutional Review Board (IRB) approval was obtained following institutional guidelines using de-identified human cancer paraffin sections. Tumor tissue paraffin sections and multiple tumor arrays were all evaluated histologically by the study pathologist before FISH and immunohistochemistry analyses. For the FISH assay, the paraffin-embedded slides were baked, deparaffinized, treated with tissue pre-treatment reagents (Insitus Biotechnologies, NM), and hybridized using Vysis reagents. A BAC clone (RP11-527H7) containing the RBM14 gene, confirmed by PCR, was labeled with Spectrum Red dUTP (Nick Translation Kit, Vysis, IL) according to the manufacturer's protocol to produce the probe for interphase chromosome hybridization. Chromosome 11 centromeric alpha satellite probe CEP11-D11Z1 (Vysis) labeled in green was applied simultaneously as a control probe for dual-color visualization. Slides were washed under standard conditions and counterstained with DAPI II before visualization with fluorescent microscopy.

Immunohistochemistry

Polyclonal rabbit antibody against GT198 was affinity purified. FFPE sections or tumor microarrays were deparaffinized and dehydrated through xylene and ethanol series, followed by antigen retrieval in 10 mM sodium citrate buffer, pH 6.0, containing 0.05% Triton at 90ºC for 20 min. Anti-GT198 (1:150) was incubated at 4ºC overnight. Antibody binding was detected using biotinylated anti-rabbit antibody and streptavidin-HRP, followed by DAB (brown) detecting reagents (Abcam, ab64261). Sections were counterstained with hematoxylin (blue).

GT198 protein purification

N-terminal His-tagged recombinant human GT198 protein (aa 1-180) was expressed in *E. coli* BL21(DE3)pLysS and purified through Ni-NTA-agarose (Qiagen, #30210). C-terminal deletion of GT198 is essential to release protein auto-inhibition *in vitro*. GT198 proteins were eluted by 200 mM imidazole, desalted and concentrated using Amicon YM-10 spin columns before use.

GT198 DNA-binding assays

The binding of biotinylated DNA to GT198 was detected by chemiluminescence. The DNA binding to GT198 is non-sequence specific. A single-stranded 25-mer biotinylated oligonucleotide [Biotin]-cctggggttgctgaggtcctggcag was used in the assay since it is sufficient to bind one GT198 dimer. White MicroLite™ 2+ 96-well plates (Thermo Scientific, #7572) were coated overnight at 37ºC with 400 ng/well of recombinant His-tagged GT198 proteins together with 5 μg/well of purified BSA (NEB) in a volume of 50 μl. BSA alone was included as a control for background subtraction. The binding in duplicates (n = 2) was carried out for 4 hrs to overnight at 4ºC. The binding was carried out with serial five-fold diluted chemical drugs (0.128, 0.64, 3.2, 16, 80, 400, 2000, 10000 nM) or

herbal ethanol extracts (0.0015, 0.007, 0.038, 0.192, 0.96, 4.8, 24, 120, 600 ng/μl). The plates were then washed four times with TPBS for a total 60 min, incubated with streptavidin-conjugated horseradish peroxidase (HRP) (Roche Molecular Biochemicals, #1089153) at 1 U/ml for 1 h at 4°C, and further washed three times with TPBS in 30 min. Bound DNA were detected by chemiluminescence with 50 μl/well ECL detection reagents (Amersham Pharmacia Biotech) in a Dynex MLX luminometer.

GT198 vaccination in MMTV-PyMT mouse tumor models

The research protocols were approved by the Institutional Animal Care and Use Committee (IACUC). Denatured human GT198 antigen for vaccination was produced from insoluble inclusion body of recombinant GST-GT198 protein isolated from E. coli BL21 (DE3), since insoluble antigen has greater efficacy in vaccination. Briefly, the isolated inclusion body containing 95% pure GT198 protein was repeatedly washed by sonication in PBS and sterilized by 70% ethanol. Incomplete Freund's adjuvant (IFA) was mixed with PBS at 1:1 ratio together with GT198 protein pellet, and was sonicated using a sterilized probe to produce GT198 antigen at 1 mg/ml for each subcutaneous injection at 100 μg in 100 μl (at the age of weeks 4, 6, 8, 10, 12, 14, 16). Mouse tail blood from MMTV-PyMT mice was collected at each vaccine time point to produce serum (5-10 μl). The antibody titers were measured at the end of experiment using His-tagged GT198-coated 96-well white plate (100 ng GT198 and 5 μg BSA/well), which was incubated with 200 μl of 1:1000 diluted mouse sera in duplicate wells, and detected by HRP-conjugated anti-mouse antibody with ECL detection reagents. Antibody titers were counted by a Dynex luminometer. Tumor sizes were measured at the each vaccine time point for analysis.

Statistical analysis

Statistical analyses were carried out using GraphPad Prism software. P values in scattergrams were calculated using unpaired two-tailed t test. * $P<0.05$, ** $P<0.01$, *** $P<0.001$; NS, not significant. A P value of less than 0.05 is considered statistically significant.

REFERENCES:

1. Sui, Y., Yang, Z., Xiong, S., Zhang, L., Blanchard, K. L., Peiper, S. C., Dynan, W. S., Tuan, D., and Ko, L. (2007) Gene amplification and associated loss of 5' regulatory sequences of CoAA in human cancers. *Oncogene* **26**, 822-835
2. Yang, Z., Peng, M., Cheng, L., Jones, K., Maihle, N. J., Mivechi, N. F., and Ko, L. (2016) GT198 Expression Defines Mutant Tumor Stroma in Human Breast Cancer. *Am J Pathol* **186**, 1340-1350
3. Zhang, L., Wang, Y., Rashid, M. H., Liu, M., Angara, K., Mivechi, N. F., Maihle, N. J., Arbab, A. S., and Ko, L. (2017) Malignant pericytes expressing GT198 give rise to tumor cells through angiogenesis. *Oncotarget* **8**, 51591-51607
4. Peng, M., Yang, Z., Zhang, H., Jaafar, L., Wang, G., Liu, M., Flores-Rozas, H., Xu, J., Mivechi, N. F., and Ko, L. (2013) GT198 splice variants display dominant negative activities and are induced by inactivating mutations. *Genes Cancer* **4**, 26-38
5. Brooks, Y. S., Wang, G., Yang, Z., Smith, K. K., Bieberich, E., and Ko, L. (2009) Functional pre- mRNA trans-splicing of coactivator CoAA and corepressor RBM4 during stem/progenitor cell differentiation. *J Biol Chem* **284**, 18033-18046
6. Peng, M., Zhang, H., Jaafar, L., Risinger, J. I., Huang, S., Mivechi, N. F., and Ko, L. (2013) Human Ovarian Cancer Stroma Contains Luteinized Theca Cells Harboring Tumor Suppressor Gene GT198 Mutations. *J Biol Chem* **288**, 33387-33397
7. Peng, M., Bakker, J. L., DiCioccio, R. A., Gille, J. J. P., Zhao, H., Odunsi, K., Sucheston, L., Jaafar, L., Mivechi, N. F., Waisfisz, Q., and Ko, L. (2013) Inactivating mutations in GT198 in familial and early-onset breast and ovarian cancers. *Genes Cancer* **4**, 15-25
8. Schubert, S., Ripperger, T., Rood, M., Petkidis, A., Hofmann, W., Frye-Boukhriss, H., Tauscher, M., Auber, B., Hille-Betz, U., Illig, T., Schlegelberger, B., and Steinemann, D. (2017) GT198 (PSMC3IP) germline variants in early-onset breast cancer patients from hereditary breast and ovarian cancer families. *Genes Cancer* **8**, 472-483

9. Breivik, J., and Gaudernack, G. (1999) Genomic instability, DNA methylation, and natural selection in colorectal carcinogenesis. *Semin Cancer Biol* **9**, 245-254
10. Albertson, D. G., Collins, C., McCormick, F., and Gray, J. W. (2003) Chromosome aberrations in solid tumors. *Nat Genet* **34**, 369-376
11. Jackson, A. L., and Loeb, L. A. (1998) On the origin of multiple mutations in human cancers. *Semin Cancer Biol* **8**, 421-429
12. Stark, G. R., and Wahl, G. M. (1984) Gene amplification. *Annu Rev Biochem* **53**, 447-491
13. Aman, P. (1999) Fusion genes in solid tumors. *Semin Cancer Biol* **9**, 303-318
14. Sorensen, P. H., and Triche, T. J. (1996) Gene fusions encoding chimaeric transcription factors in solid tumours. *Semin Cancer Biol* **7**, 3-14
15. Rabbitts, T. H. (1994) Chromosomal translocations in human cancer. *Nature* **372**, 143-149
16. Vogelstein, B., and Kinzler, K. W. (1994) Has the breast cancer gene been found? *Cell* **79**, 1-3
17. Koreth, J., Bakkenist, C. J., and McGee, J. O. (1999) Chromosomes, 11Q and cancer: a review. *J Pathol* **187**, 28-38
18. Ronchetti, D., Finelli, P., Richelda, R., Baldini, L., Rocchi, M., Viggiano, L., Cuneo, A., Bogni, S., Fabris, S., Lombardi, L., Maiolo, A. T., and Neri, A. (1999) Molecular analysis of 11q13 breakpoints in multiple myeloma. *Blood* **93**, 1330-1337
19. Chandrasekharappa, S. C., Guru, S. C., Manickam, P., Olufemi, S. E., Collins, F. S., Emmert-Buck, M. R., Debelenko, L. V., Zhuang, Z., Lubensky, I. A., Liotta, L. A., Crabtree, J. S., Wang, Y., Roe, B. A., Weisemann, J., Boguski, M. S., Agarwal, S. K., Kester, M. B., Kim, Y. S., Heppner, C., Dong, Q., Spiegel, A. M., Burns, A. L., and Marx, S. J. (1997) Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* **276**, 404-407
20. Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., and King, M. C. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-1689
21. Narod, S. A., Feunteun, J., Lynch, H. T., Watson, P., Conway, T., Lynch, J., and Lenoir, G. M. (1991) Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet* **338**, 82-83
22. Rommens, J. M., Durocher, F., McArthur, J., Tonin, P., LeBlanc, J. F., Allen, T., Samson, C., Ferri, L., Narod, S., Morgan, K., and et al. (1995) Generation of a transcription map at the HSD17B locus centromeric to BRCA1 at 17q21. *Genomics* **28**, 530-542
23. King, M. C. (2014) "The race" to clone BRCA1. *Science* **343**, 1462-1465
24. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., and et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71
25. Ko, L., Cardona, G. R., Henrion-Caude, A., and Chin, W. W. (2002) Identification and characterization of a tissue-specific coactivator, GT198, that interacts with the DNA-binding domains of nuclear receptors. *Mol Cell Biol* **22**, 357-369
26. Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M. D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T. R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B. A., Gayther, S. A., Zelada-Hedman, M., and et al. (1998) Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* **62**, 676-689
27. Iwasaki, T., Chin, W. W., and Ko, L. (2001) Identification and characterization of RRM-containing coactivator activator (CoAA) as TRBP-interacting protein, and its splice variant as a coactivator modulator (CoAM). *J Biol Chem* **276**, 33375-33383
28. Perani, M., Antonson, P., Hamoudi, R., Ingram, C. J., Cooper, C. S., Garrett, M. D., and Goodwin, G. H. (2005) The proto-oncoprotein SYT interacts with SYT-interacting protein/co-activator activator (SIP/CoAA), a human nuclear receptor co-activator with similarity to EWS and TLS/FUS family of proteins. *J Biol Chem* **280**, 42863-42876

29. Liu, P., Greenleaf, A. L., and Stiller, J. W. (2008) The Essential Sequence Elements Required for RNAP II C-terminal Domain Function in Yeast and their Evolutionary Conservation. *Mol Biol Evol* **25**, 719-727
30. Stiller, J. W., and Cook, M. S. (2004) Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell* **3**, 735-740
31. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-1876
32. Ko, L., Xiong, S., Xu, W., and O'Malley, B. W. (2008) Tyrosine-rich repeats of coactivator CoAA interact with the CTD of RNA Polymerase II. *ASBMB Annual Meeting of Transcription Regulation by Chromatin and RNA Polymerase II*, Granlibakken, CA.
33. Yang, L., Chansky, H. A., and Hickstein, D. D. (2000) EWS.Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. *J Biol Chem* **275**, 37612-37618
34. Bennett, C. B., Westmoreland, T. J., Verrier, C. S., Blanchette, C. A., Sabin, T. L., Phatnani, H. P., Mishina, Y. V., Huper, G., Selim, A. L., Madison, E. R., Bailey, D. D., Falae, A. I., Galli, A., Olson, J. A., Greenleaf, A. L., and Marks, J. R. (2008) Yeast screens identify the RNA polymerase II CTD and SPT5 as relevant targets of BRCA1 interaction. *PLoS ONE* **3**, e1448
35. Yang, Z., Sui, Y., Xiong, S., Liour, S. S., Phillips, A. C., and Ko, L. (2007) Switched Alternative Splicing of Oncogene CoAA during embryonal carcinoma stem cell differentiation. *Nucleic Acids Res* **35**, 1919-1932
36. Jackson, F. R., Banfi, S., Guffanti, A., and Rossi, E. (1997) A novel zinc finger-containing RNA-binding protein conserved from fruitflies to humans. *Genomics* **41**, 444-452
37. Rabbitts, T. H. (1999) Perspective: chromosomal translocations can affect genes controlling gene expression and differentiation--why are these functions targeted? *J Pathol* **187**, 39-42
38. Sanford, J. R., and Caceres, J. F. (2004) Pre-mRNA splicing: life at the centre of the central dogma. *J Cell Sci* **117**, 6261-6263
39. Smith, C. W., Patton, J. G., and Nadal-Ginard, B. (1989) Alternative splicing in the control of gene expression. *Annu Rev Genet* **23**, 527-577
40. Venables, J. P. (2006) Unbalanced alternative splicing and its significance in cancer. *Bioessays* **28**, 378-386
41. Graveley, B. R. (2011) Splicing up pluripotency. *Cell* **147**, 22-24
42. Morrison, S. J., and Kimble, J. (2006) Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068-1074
43. Khoury, M. P., and Bourdon, J. C. (2011) p53 Isoforms: An Intracellular Microprocessor? *Genes Cancer* **2**, 453-465
44. Orban, T. I., and Olah, E. (2003) Emerging roles of BRCA1 alternative splicing. *Mol Pathol* **56**, 191-197
45. Auboeuf, D., Dowhan, D. H., Li, X., Larkin, K., Ko, L., Berget, S. M., and O'Malley, B. W. (2004) CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. *Mol Cell Biol* **24**, 442-453
46. Enomoto, R., Kinebuchi, T., Sato, M., Yagi, H., Shibata, T., Kurumizaka, H., and Yokoyama, S. (2004) Positive role of the mammalian TBPIP/HOP2 protein in DMC1-mediated homologous pairing. *J Biol Chem* **279**, 35263-35272
47. Pezza, R. J., Voloshin, O. N., Vanevski, F., and Camerini-Otero, R. D. (2007) Hop2/Mnd1 acts on two critical steps in Dmc1-promoted homologous pairing. *Genes Dev* **21**, 1758-1766
48. Sansam, C. L., and Pezza, R. J. (2015) Connecting by breaking and repairing: mechanisms of DNA strand exchange in meiotic recombination. *Febs J* **282**, 2444-2457
49. Chi, P., San Filippo, J., Sehorn, M. G., Petukhova, G. V., and Sung, P. (2007) Bipartite stimulatory action of the Hop2-Mnd1 complex on the Rad51 recombinase. *Genes Dev* **21**, 1747-1757
50. Cho, N. W., Dilley, R. L., Lampson, M. A., and Greenberg, R. A. (2014) Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell* **159**, 108-121

51. Wall, M. E., and Wani, M. C. (1995) Camptothecin and taxol: discovery to clinic--thirteenth Bruce F. Cain Memorial Award Lecture. *Cancer Res* **55**, 753-760
52. Wani, M. C., and Horwitz, S. B. (2014) Nature as a remarkable chemist: a personal story of the discovery and development of Taxol. *Anticancer Drugs* **25**, 482-487
53. Komlodi-Pasztor, E., Sackett, D. L., and Fojo, T. (2012) Tales of how great drugs were brought down by a flawed rationale. *Clin Cancer Res* **19**, 1304
54. Parness, J., and Horwitz, S. B. (1981) Taxol binds to polymerized tubulin in vitro. *J Cell Biol* **91**, 479-487
55. Ozcelik, B., Turkyilmaz, C., Ozgun, M. T., Serin, I. S., Batukan, C., Ozdamar, S., and Ozturk, A. (2010) Prevention of paclitaxel and cisplatin induced ovarian damage in rats by a gonadotropin-releasing hormone agonist. *Fertil Steril* **93**, 1609-1614
56. Kadota, T., Chikazawa, H., Kondoh, H., Ishikawa, K., Kawano, S., Kuroyanagi, K., Hattori, N., Sakakura, K., Koizumi, S., Hiraiwa, E., and et al. (1994) Toxicity studies of paclitaxel. (I)--Single dose intravenous toxicity in rats. *J Toxicol Sci* **19 Suppl 1**, 1-9
57. Poruchynsky, M. S., Komlodi-Pasztor, E., Trostel, S., Wilkerson, J., Regairaz, M., Pommier, Y., Zhang, X., Kumar Maity, T., Robey, R., Burotto, M., Sackett, D., Guha, U., and Fojo, A. T. (2015) Microtubule-targeting agents augment the toxicity of DNA-damaging agents by disrupting intracellular trafficking of DNA repair proteins. *Proc Natl Acad Sci U S A* **112**, 1571-1576
58. Miller, L. H., and Su, X. (2011) Artemisinin: discovery from the Chinese herbal garden. *Cell* **146**, 855-858
59. Shoemaker, M., Hamilton, B., Dairkee, S. H., Cohen, I., and Campbell, M. J. (2005) In vitro anticancer activity of twelve Chinese medicinal herbs. *Phytother Res* **19**, 649-651
60. Ryu, S., Park, K. M., and Lee, S. H. (2016) Gleditsia sinensis Thorn Attenuates the Collagen-Based Migration of PC3 Prostate Cancer Cells through the Suppression of alpha2beta1 Integrin Expression. *Int J Mol Sci* **17**, 328
61. Lee, S. J., Park, K., Ha, S. D., Kim, W. J., and Moon, S. K. (2010) Gleditsia sinensis thorn extract inhibits human colon cancer cells: the role of ERK1/2, G2/M-phase cell cycle arrest and p53 expression. *Phytother Res* **24**, 1870-1876
62. Lu, D., Xia, Y., Tong, B., Zhang, C., Pan, R., Xu, H., Yang, X., and Dai, Y. (2014) In vitro anti-angiogenesis effects and active constituents of the saponin fraction from Gleditsia sinensis. *Integr Cancer Ther* **13**, 446-457
63. Lee, J., Yi, J. M., Kim, H., Lee, Y. J., Park, J. S., Bang, O. S., and Kim, N. S. (2014) Cytochalasin H, an active anti-angiogenic constituent of the ethanol extract of Gleditsia sinensis thorns. *Biol Pharm Bull* **37**, 6-12
64. Yi, J. M., Park, J. S., Oh, S. M., Lee, J., Kim, J., Oh, D. S., Bang, O. S., and Kim, N. S. (2012) Ethanol extract of Gleditsia sinensis thorn suppresses angiogenesis in vitro and in vivo. *BMC Complement Altern Med* **12**, 243
65. Gao, J., Yang, X., and Yin, W. (2016) From Traditional Usage to Pharmacological Evidence: A Systematic Mini-Review of Spina Gleditsiae. *Evid Based Complement Alternat Med* **2016**, 3898957
66. Shamaladevi, N., Lyn, D. A., Shaaban, K. A., Zhang, L., Villate, S., Rohr, J., and Lokeshwar, B. L. (2013) Ericifolin: a novel antitumor compound from allspice that silences androgen receptor in prostate cancer. *Carcinogenesis* **34**, 1822-1832
67. Chevallier, A. (2016) *Encyclopedia of Herbal Medicine*, Dorling Kindersley Limited
68. Oberlies, N. H., and Kroll, D. J. (2004) Camptothecin and taxol: historic achievements in natural products research. *J Nat Prod* **67**, 129-135
69. Schiff, P. B., and Horwitz, S. B. (1980) Taxol stabilizes microtubules in mouse fibroblast cells. *Proc Natl Acad Sci U S A* **77**, 1561-1565
70. Csiszar, A. (2016) Peer review: Troubled from the start. *Nature* **532**, 306-308
71. Chodosh, L. A. (1998) Expression of BRCA1 and BRCA2 in normal and neoplastic cells. *J Mammary Gland Biol Neoplasia* **3**, 389-402