

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2017 November ; 26(11): 1667–1675. doi:
10.1158/1055-9965.EPI-17-0232.

Developing the WCRF International/University of Bristol methodology for identifying and carrying out systematic reviews of mechanisms of exposure-cancer associations

Sarah J Lewis^{1,2}, Mike Gardner^{1,3}, Julian Higgins¹, Jeff MP Holly⁴, Tom R Gaunt^{1,2}, Claire M Perks⁴, Suzanne D Turner⁵, Sabina Rinaldi⁶, Steve Thomas⁷, Sean Harrison^{1,2}, Rosie J Lennon¹, Vanessa Tan^{1,2}, Cath Borwick^{1,8}, Pauline Emmett¹, Mona Jeffreys¹, Kate Northstone⁹, Giota Mitrou¹⁰, Martin Wiseman¹⁰, Rachel Thompson¹⁰, and Richard M Martin^{1,2,11}

¹School of Social and Community Medicine, University of Bristol, UK

²MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, UK

³Nuffield Department of Population Health, University of Oxford, UK

⁴IGFs & Metabolic Endocrinology Group, Learning and Research Building, School of Clinical Sciences, Southmead Hospital, Bristol, UK

⁵Department of Pathology, University of Cambridge, UK

⁶International Agency for Research on Cancer, Lyon, France

⁷School of Oral and Dental Sciences, University of Bristol, UK

⁸Cardiff University, Cardiff, UK

⁹CLAHRC West, University of Bristol, UK

¹⁰WCRF International, London, UK

¹¹National Institute for Health Research Biomedical Research Unit in Nutrition, Diet and Lifestyle at University Hospitals Bristol NHS Foundation Trust and the University of Bristol, BS2 8AE, Bristol, UK

Abstract

Background—Human, animal and cell experimental studies, human biomarker studies and genetic studies complement epidemiological findings and can offer insights into biological plausibility and pathways between exposure and disease but methods for synthesising such studies are lacking. We therefore developed a methodology for identifying mechanisms and carrying out systematic reviews of mechanistic studies which underpin exposure-cancer associations.

Methods—A multidisciplinary team with expertise in informatics, statistics, epidemiology, systematic reviews, cancer biology and nutrition was assembled. Five one-day workshops were

held to brainstorm ideas, in the intervening periods we carried-out searches and applied our methods to a case study to test our ideas.

Results—We have developed a two stage framework, the first stage of which is designed to identify mechanisms underpinning a specific exposure-disease relationship, the second stage is a targeted systematic review of studies on a specific mechanism. As part of the methodology we also developed an online tool for text mining for mechanism prioritization (TeMMPo) and a new graph for displaying related but heterogeneous data from epidemiological studies (the albatross plot).

Conclusions—We have developed novel tools for identifying mechanisms and carrying out systematic reviews of mechanistic studies of exposure-disease relationships. In doing so we have outlined how we have overcome the challenges that we faced and provide researchers with practical guides for conducting mechanistic systematic reviews.

Impact—The above methodology and tools will allow; potential mechanisms to be identified and the strength of the evidence underlying a particular mechanism to be assessed.

Keywords

Systematic review methodology; mechanisms; cancer

Introduction

Systematic reviews offer robust methodology for identifying, appraising and synthesising studies that have addressed a common research question [1,2]. Such reviews are valuable in the synthesis of published literature relating to health care interventions, and to aetiological questions. However, reviews of observational epidemiological findings by themselves are insufficient to establish causation. Other forms of evidence are required to complement such data in order to infer the likely causality of any observed association, in particular biological plausibility [3]. There is an abundance of evidence relating to the biology underpinning the causation of disease, from studies such as human, animal and cell experimental studies, human biomarker studies and genetic association studies, although methods have not been developed to synthesise this in a systematic way. Consequently, whilst epidemiological studies addressing chronic disease can be synthesised using a systematic process, mechanistic studies have previously been addressed using a results narrative.

The World Cancer Research Fund and American Institute for Cancer Research have published a landmark report addressing the prevention of cancer through diet, nutrition and physical activity [4]. As part of the Continuous Update of the 2007 Report [5] WCRF UK commissioned the University of Bristol to develop a framework for reviewing mechanistic studies of exposures and cancer to test the likely causality of the observed associations. The aims were to: i) identify mechanistic studies that provide evidence of the biological plausibility of the causality of links between a diet, nutrition or physical activity exposure, and cancer; and ii) systematically review and assess the strength of the evidence for any one particular mechanism.

Challenges in conducting systematic reviews of the mechanisms mediating observed associations between potentially modifiable exposures and cancer

How to identify the relevant mechanisms for a particular exposure-outcome association?

How to cope with the enormous wealth of data generated in searching for mechanisms?

How to assess the quality of animal and cell studies?

How to determine the relevance of animal studies to human disease?

How to assess the extent of publication bias?

How best to integrate all the evidence?

We outline how we addressed the challenges inherent in developing an overall methodology outlined above. A schematic diagram of the steps is given in Figure 1 with full details of the methodology presented in the Supplementary material.

Materials and Methods

We approached colleagues and collaborators from the University of Bristol, University of Cambridge and the International Agency for Research on Cancer to assemble a multidisciplinary team with expertise in bioinformatics (TG), statistics (JHi, SH, KN, RM), cancer biology (JHo, CP, ST), animal studies (JHo, MG, ST), molecular biology (TG, JHo, CP, VT, ST), epidemiology (SL, PE, MJ, KN, RM), genetic epidemiology (SL, TG), nutrition (SR, PE, KN) and systematic reviews (SL, MG, JHi, RM). Our objective to develop a rigorous systematic review methodology integrating animal, cell and human studies was met through a combination of discussion workshops and advice from a panel of experts. Decisions were reached by discussion and consensus opinion and then tested in practice. Results were fed back to the team and changes were made to the methodology if needed.

We tested the framework by implementation in a case study examining the IGF pathway to determine whether this could explain observed associations between consumption of milk and incidence of prostate cancer (reported in full separately). To do this we systematically reviewed evidence on milk-intake and the IGF pathway, and between the IGF pathway and prostate cancer.[6] In this review we pooled together evidence from randomized controlled trials and other experimental studies in humans, observational, human biomarker, genetic and animal studies. The feasibility and reproducibility of our methodology has been independently tested by two teams of systematic reviewers who initially searched for mechanisms between higher body fatness and postmenopausal breast cancer, and systematically reviewed the insulin-like growth factor 1 receptor as a potential mechanism for this association.[7] The findings by Ertaylan et al are published as an article in the same issue of this Journal. [7]

Results

Identifying the relevant mechanisms for a particular exposure-outcome association

We have developed a two-stage strategy, in stage 1 all potential mechanisms underlying a particular exposure-outcome association are identified, taking a largely ‘hypothesis-free’ approach, in stage 2 the evidence underlying one or more specific mechanisms are systematically reviewed. Fundamental to our approach are ‘intermediate phenotypes’ (IPs) between the exposure and disease (e.g. measures of DNA damage) as mechanistic studies frequently have an IP rather than cancer as an outcome, or will investigate the IP as the exposure in relation to an outcome. Stage 1 assembles the evidence around IPs, to determine which have evidence linking them either to the exposure or to the outcome, and to quantify this evidence. For the study of milk and prostate cancer a list of potential IPs was generated (Table 1). In doing this we considered the biological processes that may lead to prostate cancer, referring to important reviews in the area of cancer such as those on the hallmarks of cancer, [8] which have been proposed as a framework for considering disordered biology in malignancies. In addition, reviews specific to the cancer site (in our case prostate cancer) were consulted to identify potential mechanisms. General MeSH terms relating to potential IPs were used in the search whenever possible, rather than more specific terms, as this allowed a broader search to be carried-out. Reviewers can generate their own list of IPs by listing terms relating to general cancer processes (such as the hallmarks of cancer), searching for reviews on the biology of their cancer site of interest and seeking expert opinion. We would advocate being as inclusive as possible at this point.

Coping with the enormous wealth of data which is generated in searching for mechanisms

The sheer number of papers generated in stage 1 (>39,000 in our case study of milk and prostate cancer) meant that we needed an efficient strategy for processing these data and prioritizing mechanisms for full systematic review in stage 2. Therefore, we have devised an automated process (‘Text Mining for Mechanism Prioritisation’, TeMMPo) which allows quantification and visualisation of the amount of evidence underlying each step in the mechanistic pathway ($E \rightarrow IP$, $IP \rightarrow C$, $E \rightarrow C$, where E is exposure, IP is intermediate phenotype and C is Cancer). This tool can be accessed at <https://www.temmpo.org.uk/>. The programme allows users to upload the results of their MEDLINE or PubMed searches, which are then displayed according to the intermediate phenotypes in a Sankey plot. This illustrates the quantity of evidence linking specific IPs with exposures ($E \rightarrow IP$) and the quantity of evidence linking the same IPs with disease ($IP \rightarrow C$); the relative number of publications underlying each link is depicted by the thickness of the lines linking the terms. A weighted score is generated as follows, the number of publications for E-IP or IP-C (whichever is the least) divided by the number of publications for E-IP or IP-C (whichever is the greater) multiplied by the total number of publications for each intermediate phenotype. According to this score IPs are then ranked. These data then inform the selection of specific intermediates to be investigated in Stage 2. Figure 2 shows a Sankey plot generated by TeMMPo indicating the quantity of studies linking milk with an IP and the quantity of studies linking the same IP with a prostate cancer outcome.

The limitations of this approach are: it assumes that the co-occurrence of a biological mechanism with exposure or outcome in the literature represents an association rather than simply a co-occurrence of the two terms in the same paper; it assumes the mechanisms are represented by a single mediating factor; recently identified pathways will be underrepresented in this approach as they are likely to have fewer studies; and it does not address issues of study type, quality, direction and magnitude of results.

Systematically reviewing the evidence for a particular mechanism including assessing study quality

Having identified potential mechanisms underlying a particular exposure-outcome association, stage 2 systematically reviews the evidence underlying one or more specific mechanisms. For our study of milk-prostate cancer, we chose to systematically review the IGF pathway, since our stage 1 searches indicated that on combining all related IP terms, there were more studies linking IGF intermediates (i.e. a combination of IGF-I, IGF-II, IGF-IR, IGFBP3, IGFBP1) with both milk and prostate cancer than for other potential mechanisms.

Stage 2 largely follows standard systematic review methodology (see Appendix 1): specification of research objectives; conduct searches (see Supplementary table 1 as a guide for developing search terms); apply inclusion/exclusion criteria; extract data; assess study quality and synthesise data across studies. Existing tools for assessing study quality have not been validated or established for mechanistic [9–11] nor animal studies [12]. We recommend the Cochrane risk of bias tools for human studies [9] and SYRCLE (Systematic Review Centre for Laboratory animal Experimentation)[13], which adapts the Cochrane tool[9], for aspects of bias that are specific to animal studies. SYRCLE addresses the following domains:

- Bias due to confounding (sequence generation, baseline characteristics, allocation concealment)
- Bias due to departures from intended intervention (e.g. due to lack of random housing of animals or lack of blinding)
- Bias due to missing data
- Bias in measurement of outcomes
- Bias in selection of reported results

As far as we are aware there are currently no tools for assessing the quality of cell line studies so we developed the criteria listed in Box 3 through consensus of the Framework development group which included cell biologists. Supplementary table 2 recommends variables to extract by study type at data extraction stage in order to complete the risk of bias assessments.

Criteria used for assessing the quality of cell studies

- 1) Have the cells been obtained from a validated repository that guarantees cell verification or have the cells been appropriately independently verified?2)) Have

sufficient biological and technical repeats of the experiments been conducted and were appropriate controls included?

- 3) Were different cell lines from the same cancer type used in the study? An effect observed in more than just one cell line implies the effect is important and relevant to this cancer type.
- 4) Are culture conditions comparable between different studies?
- 5) Selective reporting: are only selected results from several cell line experiments reported?
- 6) Were cell lines from different cancer types compared? This implies an important effect that is relevant more generally to cancer cells.

We recommend that questions 1-3 above are used to determine inclusion of cell studies into the review. In our study of milk-IGF-prostate cancer only a small proportion of relevant cell studies met this basic quality criteria (Figure 3). However, it is a recent requirement to provide authentication of cell lines and other quality control criteria for publication. Thus in applying these criteria we are selecting more recent studies and may be excluding high quality historical studies which were not required to provide information on the above in order to publish. Questions 4-6 can be used to assess the reproducibility of the findings from cell studies.

Synthesis of individual studies and ‘albatross plots’ for graphical representation of evidence synthesis, when meta-analysis is not appropriate

The next step is the synthesis of data from individual studies. Formal meta-analysis of comparable studies is recommended where possible and appropriate [14]. However, it is likely that mechanistic studies will be too heterogeneous (in terms of exposure and outcome definitions; different follow-up periods; different study types) to combine, and therefore some studies will only be amenable to a narrative summary of the results. We therefore developed a new method to graphically represent heterogeneous data, which we have termed ‘albatross plots’ [15]. These plots allow for the strength and direction of association to be displayed continuously, plotting p-values against the number of participants in the studies (which will give an indication of the relative power of the study) (Figure 4). Clustering of data points towards one side of the graph represents an association between exposure and outcome in that direction. In Figure 4 the majority of studies are on the right side of the graph indicating a positive association of exposure (milk and dairy products) with outcome (IGF-I). Small studies will only have low p-values if the effect size is large, whereas large studies may have low p-values even when the effect size is small.

Contour lines which indicate a specific beta-coefficient can be added to the plot to indicate (to some extent) the magnitude of association. Simple contours can be computed based on p-values and the number of participants, although it should be noted that such contours are not sufficient or appropriate to provide a precise effect estimate (as a forest plot would).

Contours can be added if the majority of data have been analysed in the same way (linear or logistic regression, or standardised mean differences), and the contour will be of the same type of effect estimate (e.g. a standardised beta coefficient for linear regression). If data

points fall along a contour (which is shaped like a bird's wing, hence 'albatross plots') then there is likely to be an association of the magnitude represented by the contour; however, this needs to be interpreted with a narrative and consideration of the individual studies in the synthesis.

We did not find any animal or cell studies which addressed the association between milk and IGF intermediates, but the 8 animal studies on IGF-prostate cancer outcomes were too varied (different experiments, on alternative aspects of the IGF pathway, in diverse animal models, with varied outcomes), to combine in a plot. Characteristics and results of these studies were tabulated (see reference [6]). A schematic diagram of the likely biological pathway generated from animal and cell line studies is another way of presenting the data.

Assessment of the strength of evidence and classification of studies according to relevance to humans

Once the synthesis of evidence has been completed, the framework requires an assessment of the strength of the body of evidence. We recommend doing this separately for human and animal studies, according to the GRADE framework [16], which has been adopted by the Cochrane Collaboration.

Whilst our remit was to design a framework which could be used to incorporate relevant evidence from any type of study, some studies were so far removed from humans that they could not inform a judgement that a particular process is operating in the human disease pathway. However, such studies could be used to assess general biological plausibility. For cancer we chose to distinguish between two types of animal models by applying the question "Has the cancer arisen *de novo* in the animal model rather than being transplanted into the animal?" This is because transplantable models represent cancers that are already highly evolved as they have adapted growth *in vitro* (in the case of cell line xenografts) or *in vivo* growth in patient-derived xenograft models (human tumour cells taken from host patient and transplanted into immunodeficient mice), and are typically of a more aggressive biological phenotype; as such they do not closely mimic most human cancers and are unlikely to give useful information about the usual process of cancer development or progression.

We recommend that only studies that closely mimic human cancers should be used to determine the strength of the evidence underlying a particular mechanistic pathway in human cancer. Other animal studies could be assessed alongside cell line studies to determine whether they provide evidence for the general biological plausibility of the proposed mechanism.

In addition to this two-tiered distinction when applying the GRADE framework, studies are assessed according to the following criteria: indirectness (this relates to the how well the study addresses the specific research question), inconsistency, imprecision and publication bias.

As we are not aware of the GRADE framework being previously applied to animal studies, the question of indirectness in particular required some consideration. We therefore developed some questions to assess this specifically for animal studies.

Assessing the indirectness of animal studies when applying the GRADE framework

- Is the exposure applied via a route which is comparable with that in humans, and a mode which addresses the research question? (*e.g. If the interest is in a food exposure, then this should be ingested by the animal model, for other exposures it may be appropriate to introduce this via an alternative route*)
- Is the level and frequency of exposure comparable with that which humans may experience after accounting for species differences in pharmacokinetics and pharmacodynamics, or is the dose justified within the study? (*much greater doses than would be possible or reasonable in humans are unlikely to reflect human exposures*)
- Is the cancer induced (i.e. by a virus, radiation, chemical agent or genetic manipulation)? (*whether or not these studies can be included will depend on the research question, but the agent used should be relevant to the human cancer*)
- Is the time at which the outcome is assessed justified? *Whether the timing of outcome assessment is relevant will depend on the outcome; e.g. if the outcome is a gene mutation then that outcome could justifiably be assessed very quickly following exposure, but if the outcome is cancer this may require much longer follow-up to produce relevant data.*
- Does the study explore mechanisms or pathways of cancer development?
- Is the outcome of assessment cancer incidence or progression rather than surrogate measures of tumour activity such as tumour size or number of tumours?
- Do the outcome measures mimic those found in humans? More specifically, does the tumour mimic the human disease in terms of the organ or tissue affected, and at the histopathological (tissue patterns, or cell surface or intracellular protein expression levels) or genetic level (are equivalent hallmark genetic lesions observed as well as gene expression profiles). Does the progression of the disease mimic the human cancer (e.g. metastasis to the same sites, vascular and stromal invasion, response to treatment)?

If the answer to one or more of these questions is no, then the individual study should be considered to offer indirect evidence; if the majority of studies in the body of evidence are considered to offer only indirect evidence then the overall GRADE assessment across these studies should be downgraded. For example we downgraded animal studies of IGF and prostate cancer because knock-out mice do not represent variation within the normal range and in some studies the outcome measured was tumour weight or volume rather than incidence.

Investigating whether publication bias is likely to have occurred

There is empirical evidence that studies with null results (no association) are less likely to be in the published literature. Null studies may also be affected by “time lag bias” or longer time to publication. Funnel plots and the Begg [17] and Egger [18] tests can be used to examine for association between effect sizes and study sizes (essentially sample size), and such an association (‘small study effect’) may reflect publication bias. However, these approaches may not be possible due to an insufficient number of similar studies with the same exposures and outcomes measured. Ioannidis and Trikalinos [19] have developed a method to test for excess statistical significance across studies on different research questions within the same domain. Domains may be defined according to a common general theme, intervention type, subject type, methodology, research environments and language of publication or combinations of these factors. The test is a comparison of the number of observed studies with statistically significant results compared against the number of expected statistically significant results amongst all meta-analyses considered in the domain. This test can be applied to assess publication bias across domains.

An alternative approach is to qualitatively assess publication bias by obtaining data on unpublished studies (e.g. by searching the grey literature and/or contacting researchers working in the field) to determine whether relevant unpublished experiments or observational studies have been carried out. It is difficult to be systematic about such investigations, but attempts should be fully reported to ensure transparency of the process. Reviewers can then compare the results of any unpublished or grey literature studies with those which have been published to determine if there are important differences in the results. This process may indicate non-, delayed or restricted (e.g. in difficult to retrieve journals) publication of null data, suggesting distortion of the mainstream literature by publication bias.

Assessing the strength of evidence across evidence streams and synthesis of cell line and other animal studies

In the WCRF International/University of Bristol framework (Supplementary material), we have set out a model for assessing the totality of evidence by determining the strength of the overall evidence from human and animal studies which reflect the human disease process (see figure 5). In addition, we advocate using other studies to illustrate biological plausibility and illustrate the potential intricacies of the biological pathway.

Discussion

We have developed methodology which can be used to identify potential mechanisms underlying observed associations between an exposure and an outcome and to systematically review a mechanistic pathway of interest. We have overcome several hurdles including: developing an automated online tool (<https://www.temmpo.org.uk/>) to deal with the vast amounts of studies identified in stage 1; recommending tools for assessing the quality and relevance of animal and cell studies to human disease; and developing a new method for synthesising data from a variety of study types, the albatross plot. However, implementing the methodology does have some limitations, the main one being that it is very time

consuming which may constrain its use. In addition, we have seen from our case study that many animal and cell studies do not report basic information that we recommend using to assess their quality, this is particularly true for older research findings. This means that many studies which are pertinent to the research question may not be included in the overall analysis. Furthermore there is a question mark over the relevance of animal experiments to the human situation, although we have made suggestions for assessing how relevant they may be and for weighting these studies accordingly in the overall analysis.

We believe that the methodology we have developed can be applied to the integration of mechanistic studies into systematic reviews of exposures and disease in order to aid the inference of causality, and in addition may highlight gaps in our knowledge where further studies are needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would thank the Mechanisms Protocol Development Group (Drs Andrew Dannenberg, Johanna Lampe, Henry Thompson, Steven Clinton, Stephen Hursting, Nikki Ford) and Dr Susan Higginbotham (member of Secretariat) who initiated this work and whose protocol we have referred to in developing these guidelines. We would especially like to acknowledge the input of Drs Stephen Hursting and Steven Clinton in relation to assessing the relevance of animal studies to human disease.

Financial Support: All authors received a grant from the World Cancer Research Fund (grant number: RFA 2012/620). S. Harrison is a Wellcome Trust Funded PhD student, 102432/Z/13/Z. R. M. Martin, S. J. Lewis, J. M. P. Holly, T. Gaunt, C. M. Perks are supported by a Cancer Research UK (C18281/A19169) Programme Grant (the Integrative Cancer Epidemiology Programme)

References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine*. 2009; 6(6):e100097.
2. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000; 283(15): 2008–12. [PubMed: 10789670]
3. Hill AB. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*. 1965; 58(5):295–300. [PubMed: 14283879]
4. World Cancer Research Fund / American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR; 2007.
5. www.wcrf.org/cup
6. Harrison S, Lennon R, Holly J, Higgins JPT, Perks C, Gardner M, Gaunt T, Tan V, Borwick C, Emmet P, Jeffreys M, et al. Does milk intake promote prostate cancer initiation or progression via effects on insulin-like growth factors (IGFs)? A systematic review and meta-analysis. *Cancer Causes and Control*. 2017; 28:497–528. [PubMed: 28361446]
7. Ertaylan G, Le Cornet C, van Roekel EH, Jung AY, Bours MJL, Damms Machado A, van den Brandt PA, Schock H, de Kok TM, Theys J, Arts ICW, et al. A comparative study on the WCRF International/University of Bristol methodology for systematic reviews of mechanisms underpinning exposure-cancer associations. *CEBP*. In Press.

8. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144:646–674. [PubMed: 21376230]
9. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA, Cochrane Bias Methods Group. et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011 Oct 18.343:d5928. [PubMed: 22008217]
10. Sterne JA, Hernán MA, Reeves BC, Savovi J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016 Oct 12.355:i4919. [PubMed: 27733354]
11. Whiting P, Savovi J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R, the ROBIS group. ROBIS: a new tool to assess the risk of bias in systematic reviews. *Journal of Clinical Epidemiology*. 2016; 69:225–34. [PubMed: 26092286]
12. Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. *Environ Health Perspect*. 2013; 121(9): 985–92. [PubMed: 23771496]
13. Hooijmans CR, Rovers MM, de Vries RBM, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. *BMC Medical Research Methodology*. 2014; 14:43. [PubMed: 24667063]
14. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986; 7:177–188.15. [PubMed: 3802833]
15. Harrison S, Jones HE, Martin RM, Lewis S, Higgins JPT. The albatross plot: A novel graphical tool for presenting results of diversely reported studies in a systematic review. *Research Synthesis Methods*. 2017; doi: 10.1002/jrsm.1239
16. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008; 336:924–926. [PubMed: 18436948]
17. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994; 50:1088–101. [PubMed: 7786990]
18. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ*. 1997; 315:629–34. [PubMed: 9310563]
19. Ioannidis JPA, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials*. 2007; 4:245–53. [PubMed: 17715249]
20. National Toxicology Program, U.S. Department of Health and Human Services. Draft OHAT approach for systematic review and evidence integration for literature-based health assessments. 2013 Feb.

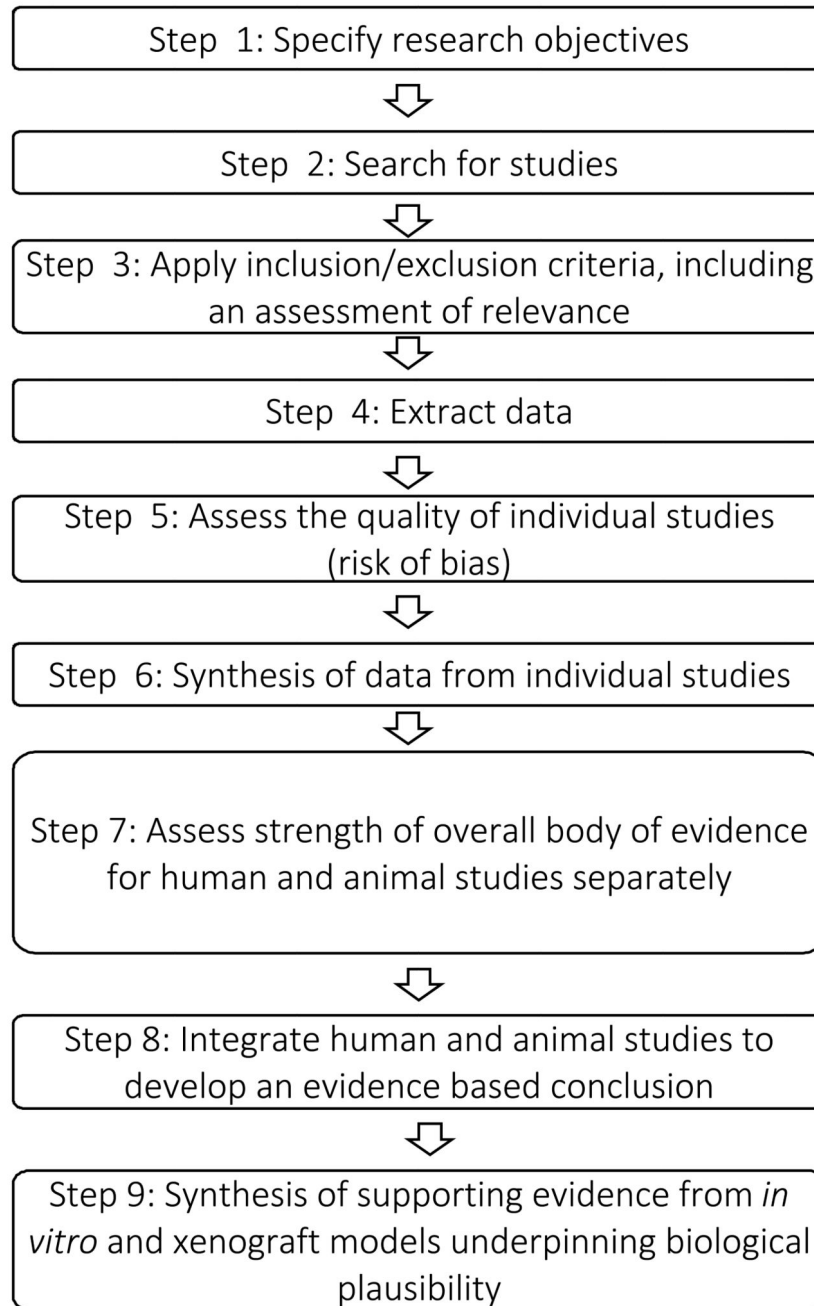


Figure 1.

Steps for stage 2

Figure 1 shows an outline of the steps we recommend going through in stage 2 of our methodology to review the evidence for a specific mechanism.

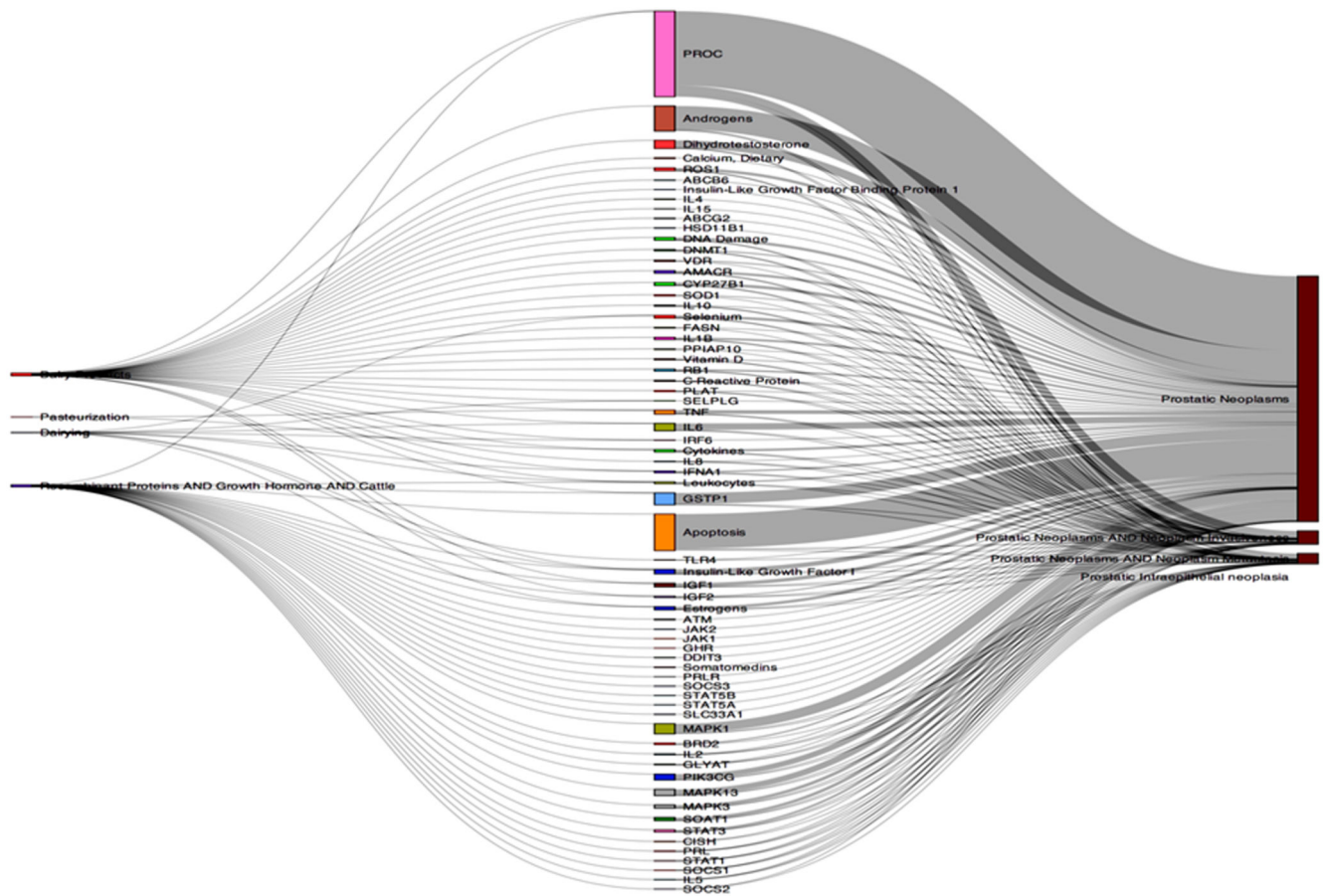


Figure 2.

A Sankey plot of milk-IGF-prostate cancer

Figure 2 shows a Sankey plot which indicates visually the quantity of evidence linking exposure to different intermediate phenotypes and the quantity of evidence linking the same intermediate phenotypes to outcome. This particular Sankey plot shows the quantity of evidence for milk and IGF on the left hand side and the quantity of evidence for IGF-prostate cancer on the right hand side of the plot.

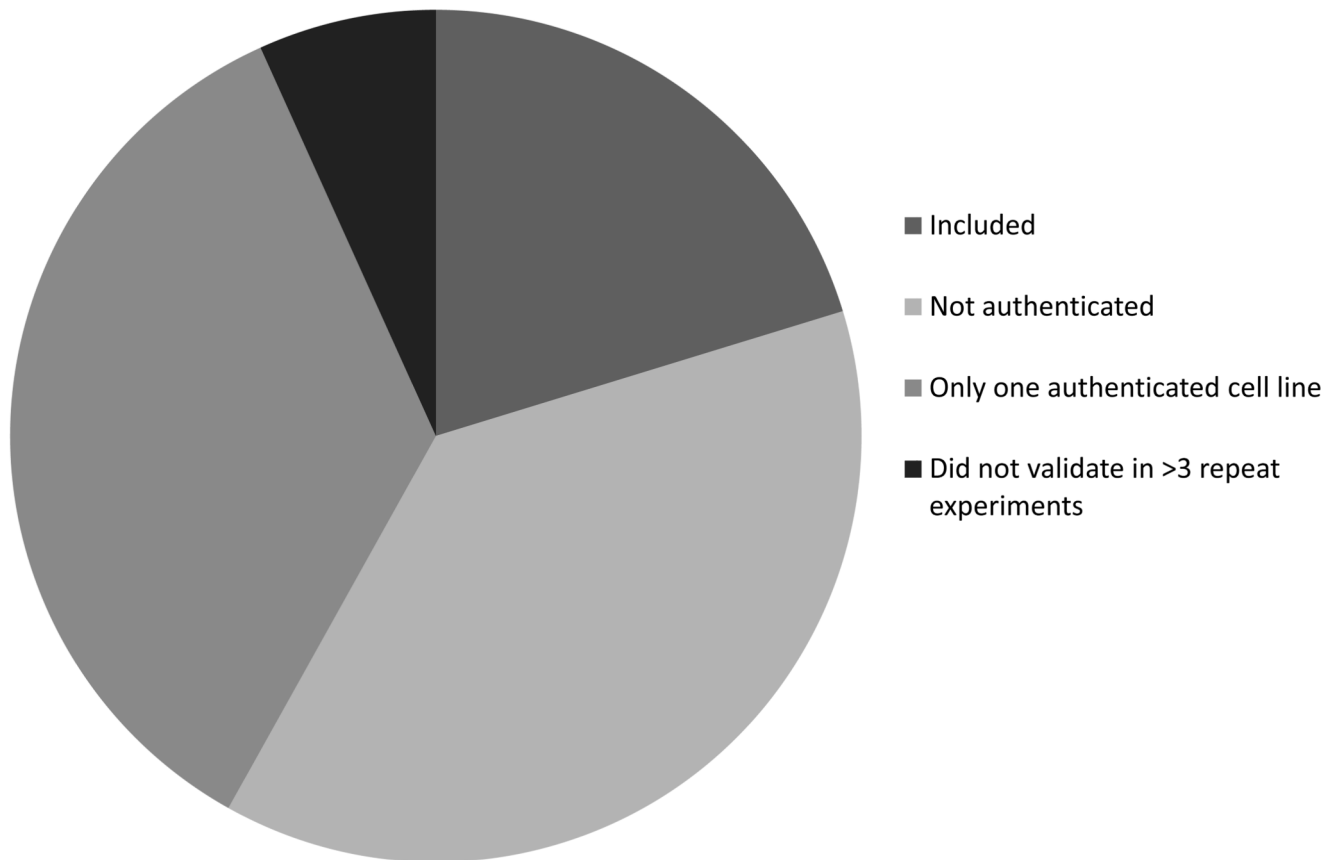


Figure 3.

Pie chart showing proportion of cell studies included after applying quality control criteria and reasons for exclusion in our study of milk-IGF-prostate cancer.

Figure 3 shows that our search identified 74 papers of cell studies relevant to milk-IGF-prostate cancer the research question; of these, 59 were excluded because they did not use authenticated cell lines (n=28); carried out experiments in only one authenticated cell line (n=26); or did not validate results in more than 3 repeat experiments (n=5).

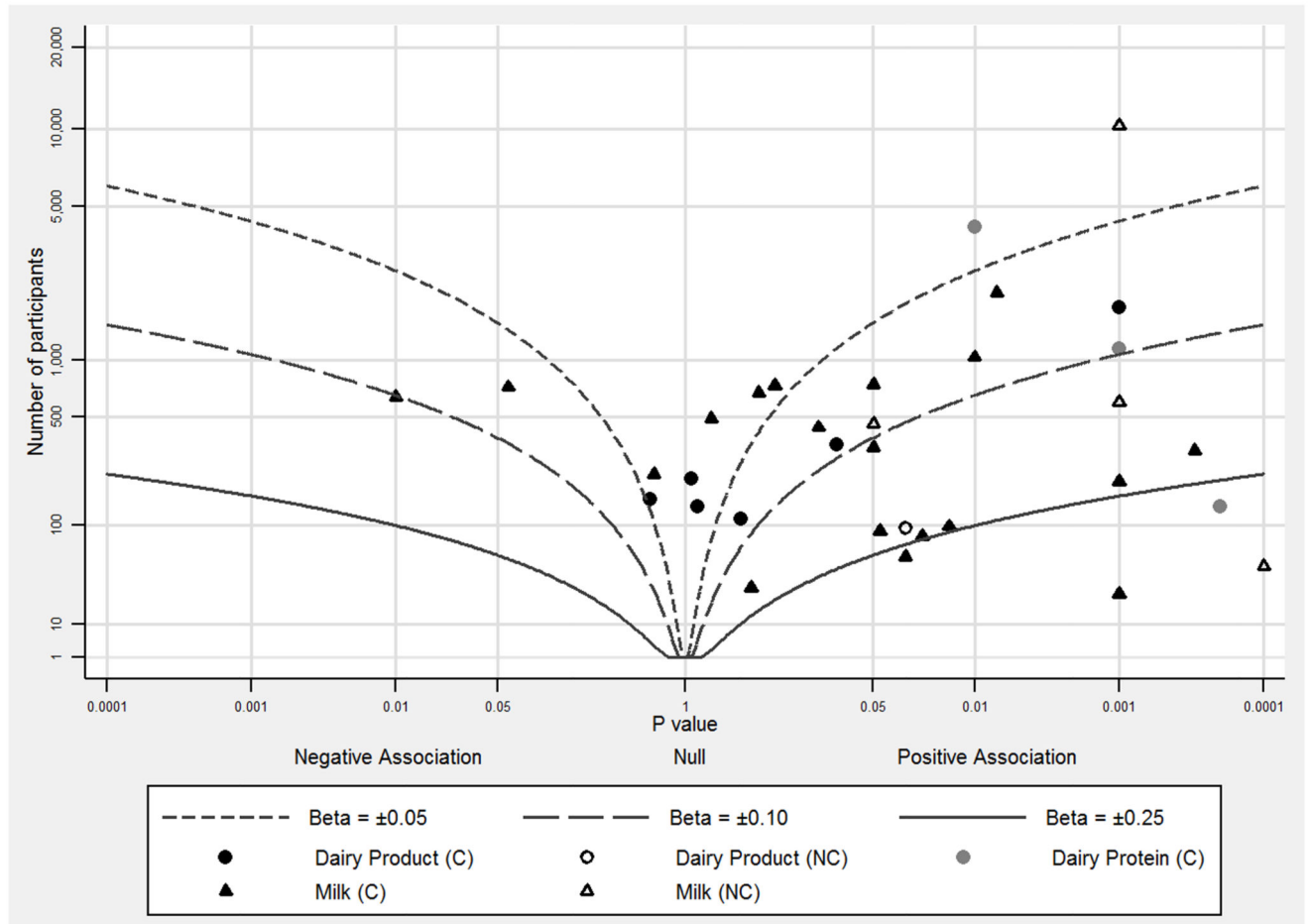
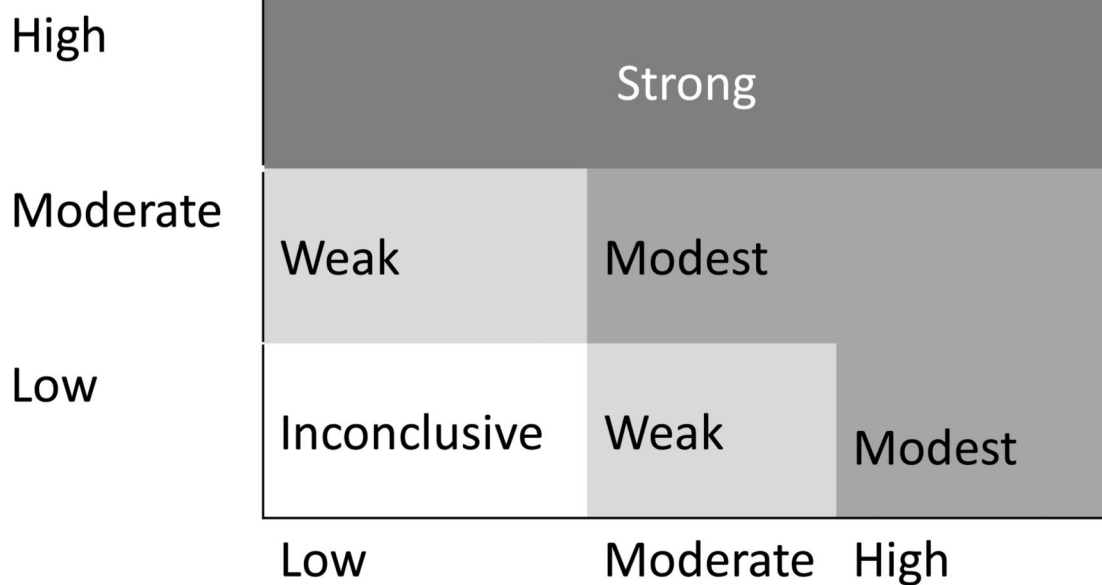


Figure 4.

Albatross plot of milk, dairy products and dairy proteins (exposures) and IGF-I (outcome). Figure 4 shows that the majority of studies are on the right side of the graph, indicating a positive association of exposure with outcome. Note also that the majority of studies showing an association do so around a standardised beta coefficient (Beta) of 0.1, which is a 0.1 standard deviation increase in outcome for a 1 standard deviation increase in exposure.

Integrating evidence

Level of evidence in human studies



Level of evidence in animal studies

Figure 5.

A guide to integrating the evidence from human and animal studies to reach an overall conclusion on the strength of evidence for a particular mechanism underlying an exposure and cancer association

Figure 5 shows how overall conclusion on the strength of evidence for exposure – intermediate and intermediate-outcome may be reached based on evidence from animal and human studies. This was adapted from the National Toxicology Program)[20]

Table 1
Intermediate phenotypes used in a review of milk and prostate cancer

MESH Terms (in bold) and more specific terms (non-bold)	Receptors, Steroid
Nerve Growth Factors	Bone Marrow
Brain-Derived Neurotrophic Factor	Enterochromaffin Cells
Ciliary Neurotrophic Factor	Immunological Synapses
Glia Maturation Factor	Leukocytes
Glial Cell Line-Derived Neurotrophic Factors	Lymphatic System
Nerve Growth Factor	Mast Cells
Neuregulins	Phagocytes
Neurotrophin 3	Mononuclear Phagocyte System
Pituitary Adenylate Cyclase-Activating Polypeptide	Angiogenesis Modulating Agents
Membrane Transport Proteins	Angiogenesis Inducing Agents
ATP-Binding Cassette Transporters	Angiogenesis Inhibitors
Amino Acid Transport Systems	Signal Transduction
Fatty Acid Transport Proteins	Ion Channel Gating
Ion Channels	Light Signal Transduction
Ion Pumps	MAP Kinase Signaling System
Monosaccharide Transport Proteins	Mechanotransduction, Cellular
Neurotransmitter Transport Proteins	Second Messenger Systems
Nucleobase, Nucleoside, Nucleotide, and Nucleic Acid Transport Proteins	Synaptic Transmission
Nucleocytoplasmic Transport Proteins	Energy Metabolism
Racemases and Epimerases	Basal Metabolism
Amino Acid Isomerases- Alanine Racemase	Citric Acid Cycle
Carbohydrate Epimerases- UDPglucose 4-Epimerase	Glycolysis
Glutathione Transferase	Oxidation-Reduction
Glutathione S-Transferase pi	Oxidative Phosphorylation
Androgens	Pentose Phosphate Pathway
Dihydrotestosterone	Photophosphorylation
Nandrolone	Proton-Motive Force
Oxandrolone	Substrate Cycling
Oxymetholone	Cell Differentiation
Stanozolol	Adipogenesis
Testosterone	Asymmetric Cell Division
Androgen Antagonists	Embryonic Induction
Chlormadinone Acetate	Gametogenesis
Cyproterone	Hematopoiesis
Cyproterone Acetate	Neurogenesis
Flutamide	Cell Death
Trans Activators	Apoptosis
Gene Products, tat	Autophagy
Herpes Simplex Virus Protein Vmw65	Necrosis

MESH Terms (in bold) and more specific terms (non-bold)	
Very Broad/General MESH terms not sub-divided for more specific terms	
Receptors, Androgen	
Receptors, Estrogen	Molecular Mechanisms
Receptors, Glucocorticoid	Physiology
Receptors, Mineralocorticoid	Cell Physiological Processes
Receptors, Progesterone	
Genomic Instability	MESH terms without more specific terms
Chromosomal Instability- Chromosome Fragility	Selenium
Microsatellite Instability	MicroRNAs
DNA Damage	DNA methylation
DNA Adducts	C-Reactive Protein
DNA Breaks- Chromosome Breakage	Telomerase
DNA Degradation, Necrotic	
DNA Fragmentation	Hormones and Growth Factors (Title- not MESH term)
DNA Repair	Testosterone
DNA End-Joining Repair	Estrogens
DNA Mismatch Repair	Somatomedins
Recombinational DNA Repair	Insulin-Like Growth Factor I
SOS Response	Insulin-Like Growth Factor II
Gene Expression	Insulin-Like Growth Factor Binding Proteins
Protein Biosynthesis	Insulin-Like Growth Factor Binding Protein 1
Transcription, Genetic- Reverse Transcription; Transcriptome	Insulin-Like Growth Factor Binding Protein 2
Mutation	Insulin-Like Growth Factor Binding Protein 3
Allelic Imbalance	Insulin-Like Growth Factor Binding Protein 4
Base Pair Mismatch	Insulin-Like Growth Factor Binding Protein 5
Chromosome Aberrations	Insulin-Like Growth Factor Binding Protein 6
Codon, Nonsense	
DNA Repeat Expansion	Vitamins and Minerals (Title- not MESH term)
Frameshift Mutation	Calcium, Dietary
Gene Amplification	Vitamin D
Gene Duplication	Mutagenesis
Germ-Line Mutation	Amino Acid SubstitutionSequence Inversion
INDEL Mutation	Chromosome Duplication
Mutagenesis, Insertional	Nondisjunction, Genetic
Mutation Rate	Somatic Hypermutation, Immunoglobulin
Mutation, Missense	Translocation, Genetic
Point Mutation	Genomic Instability
Sequence Deletion	Chromosomal Instability- Chromosome Fragility
	Suppression, Genetic
	Microsatellite Instability
Cytokines	Terms entered as Title not MESH terms
	Inflammation

MESH Terms (in bold) and more specific terms (non-bold)	Receptors, Steroid
Chemokines	Immunity
Growth Differentiation Factor 15	Programmed Cell Death
Hematopoietic Cell Growth Factors	Physiology Programmed Cell Death
Hepatocyte Growth Factor	ProstatitisPhysiology
Interferons	Physiology Prostatitis
Interleukin 1 Receptor Antagonist Protein	ProstatitisPhysiology
Interleukins	Prostatitis
Leukemia Inhibitory Factor	
Lymphokines	
Monokines	
Oncostatin M	
Osteopontin	
Transforming Growth Factor beta	
Tumor Necrosis Factors	
Cell Proliferation	
Cell Division- Asymmetric Cell Division; Telomere Homeostasis	
Immune System	
Antibody-Producing Cells	
Antigen-Presenting Cells	