

Environmental Influences on the Epigenome: Using SWIFT Text Mining Tool to Explore the State of the Science

Katherine Pelch, PhD

Office of Health Assessment and Translation
National Institute of Environmental Health Sciences

NTP Board of Scientific Counselors Meeting
June 16, 2015





Where are these links the strongest?

Environmental Exposure



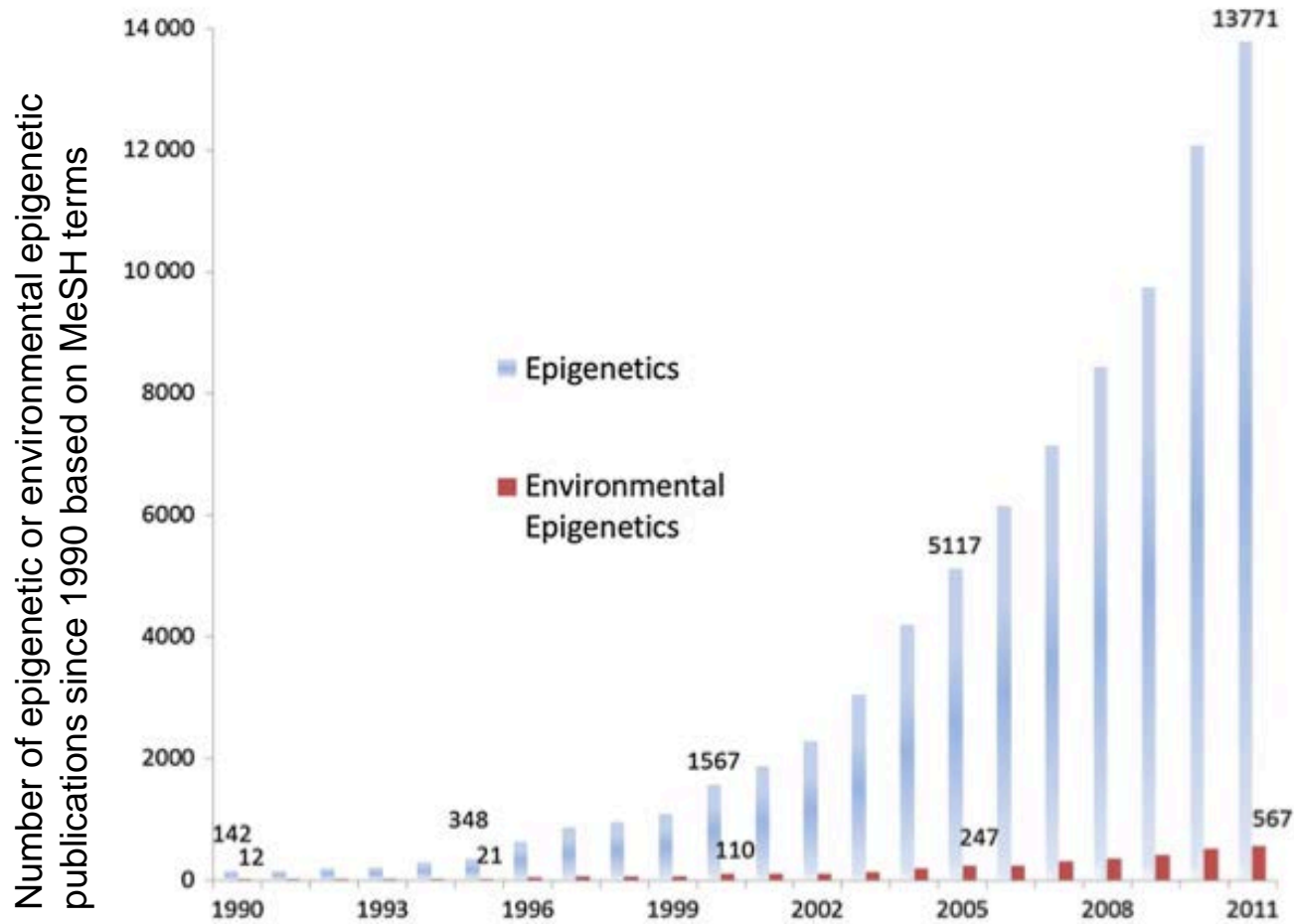
Epigenetic Mechanisms



Health Outcome



There is a growing interest in epigenetics



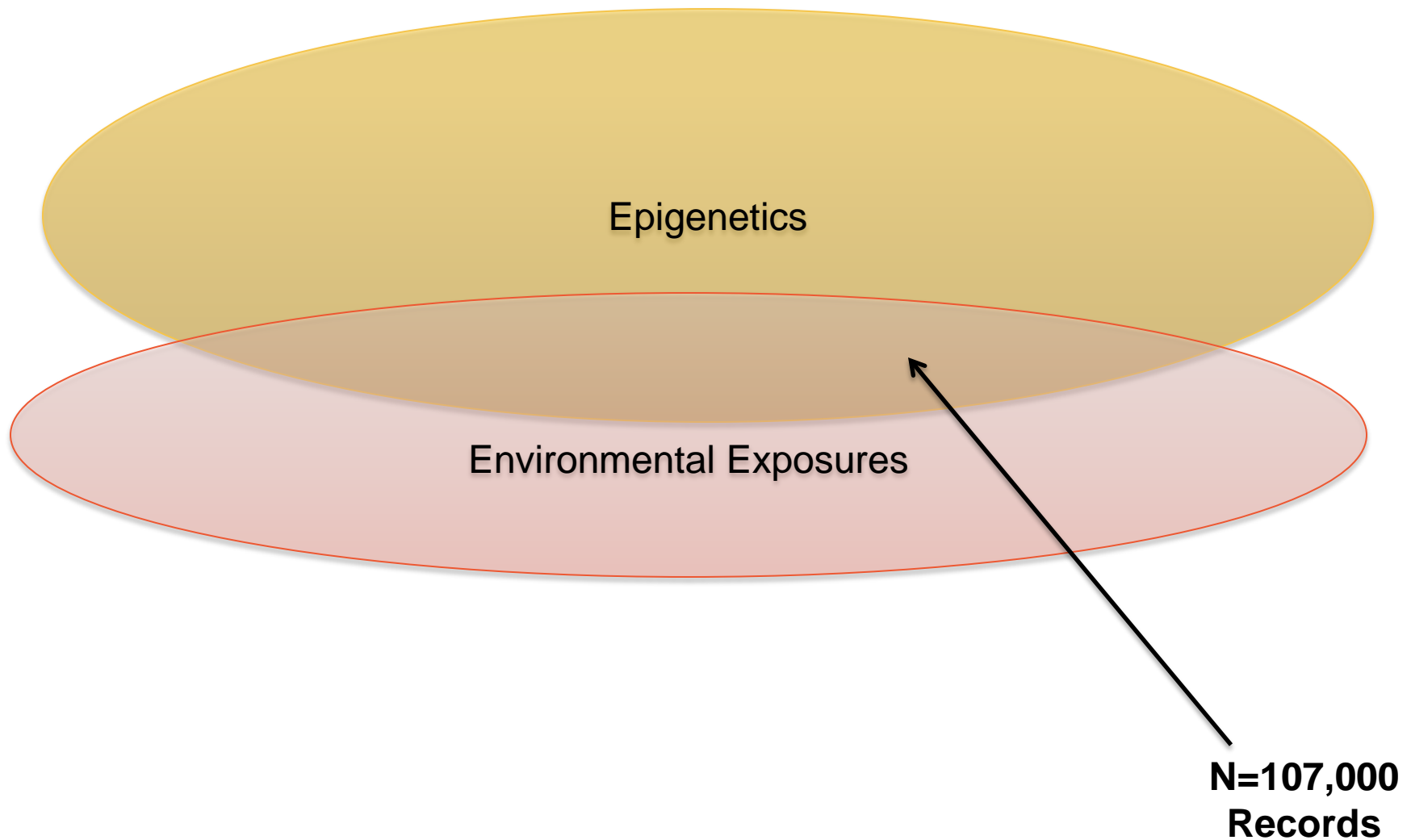
Burris HH, Baccarelli AA. 2014. Environmental epigenetics: From novelty to scientific discipline. J Appl Toxicol 34:113-116.



- 2012-2017 NIEHS Strategic Plan
 - *“the effects of the environment on the epigenetic regulation of biological and pathological processes”*
- Cross-divisional implementation group
 - *“Where are the links between environmental exposures and health outcomes via epigenetic mechanisms the strongest?”*
- Evaluation design team



How was the research question refined?



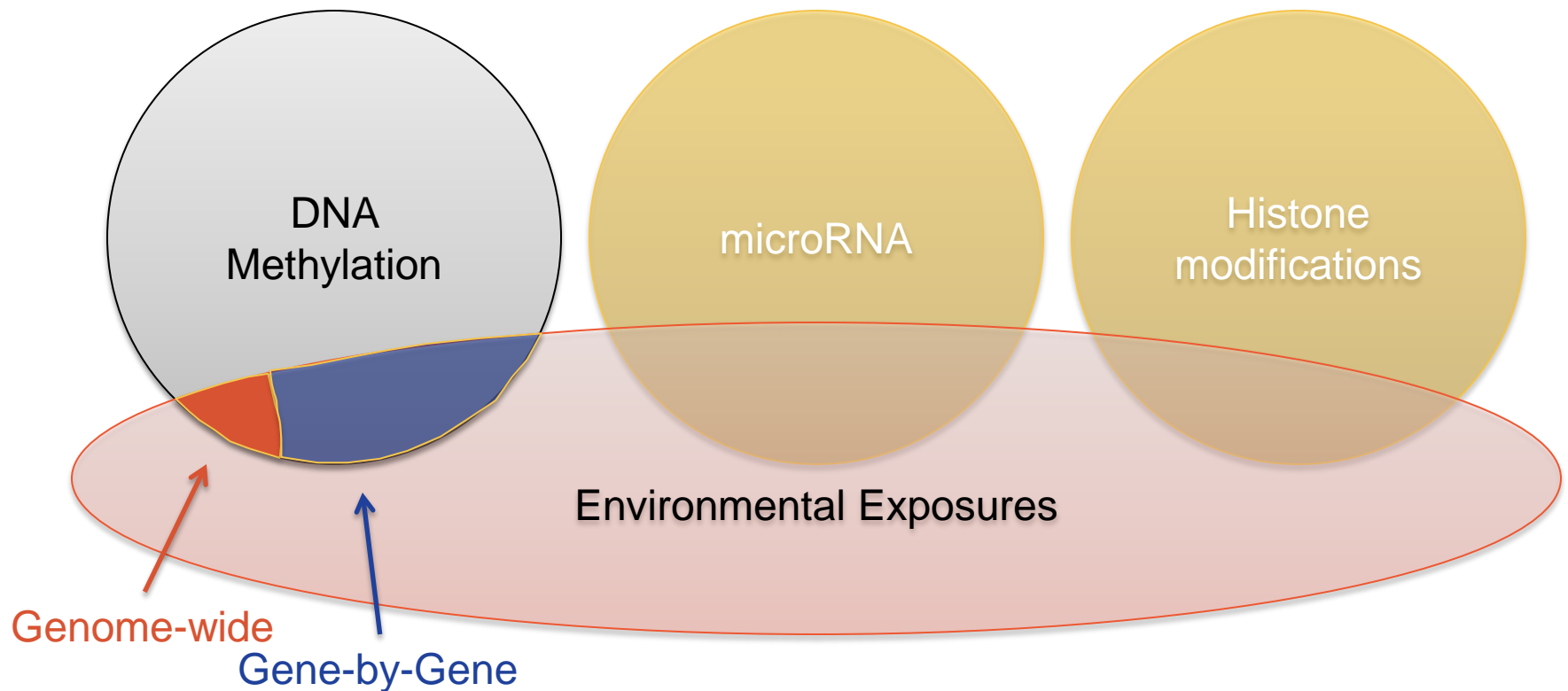


How was the research question refined?

- More focused research question
- Test drive text-mining and machine learning tools
- Develop a new report format (“scoping report”)

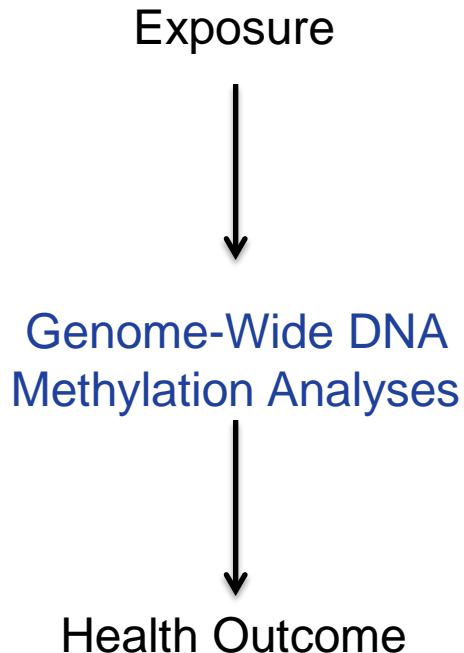


How was the research question refined?





What is the extent of the evidence?



Specific Aims

- Search for genome-wide analyses of DNA methylation
- Identify literature for seed sets
- Relevancy rank the search results
- Categorize records by type of exposure, health outcome, and evidence stream (human, animal, *in vitro*)
- Visualize the results

SCOPING REPORT



How was the literature search crafted?

- Global DNA methylation

- DNA methylation

- Genomes

- Genomes

- Epigenetics

DNA METHYLATION - Search "DNA Methylation"[mh] OR ((DNA[mh] OR DNA[tiab] OR cytosine[mh] OR cystosine*[tiab] OR Alu[tiab] OR LINE-1[tiab] OR genes[tiab] OR genom*[tiab]) AND methylat*[tiab]) OR "CpG islands"[mh] OR ((CpG[tiab] OR CpGs[tiab] OR CPG's[tiab] OR non-CG[tiab] OR non-CPG) AND methylat*[tiab]) OR differential-methylat*[tiab] OR differentially-methylated[tiab] OR partially-methylated[tiab] OR "partial methylation"[tiab] OR hemimethylat*[tiab] OR hemi-methylat*[tiab] OR hypermethylat*[tiab] OR hyper-methylat*[tiab] OR hypomethylat*[tiab] OR hypo-methylat*[tiab] OR demethylat*[tiab] OR unmethylat*[tiab] OR nonmethylat*[tiab] OR DNA modification methylases[mh] OR methylas*[tiab] OR demethylas*[tiab] OR methyltransferase*[tiab] OR "DNA Mtase"[tiab] OR Dnmt*[tiab] OR 5-methylcytosine[mh] OR "5-methylcytosine"[tiab] OR 5-methyl-cytosine[tiab] OR 5mc[tiab] OR 5meC[tiab] OR Hydroxymethylat*[tiab] OR 5-hydroxymethylcytosine[tiab] OR 5-hydroxy-methylcytosine[tiab] OR 5-hydroxymethyl-cytosine[tiab] OR 5hmC[tiab] OR Methyl-binding-domain*[tiab] OR methyl-binding-protein*[tiab] OR MBD2*[tiab] OR MBD3*[tiab] OR MBD4*[tiab] OR "methyl-cpg-binding protein 2"[mh] OR methyl-CpG*[tiab] OR MeCP2[tiab]

Global DNA methylation OR

(Epigenetics OR DNA methylation) AND

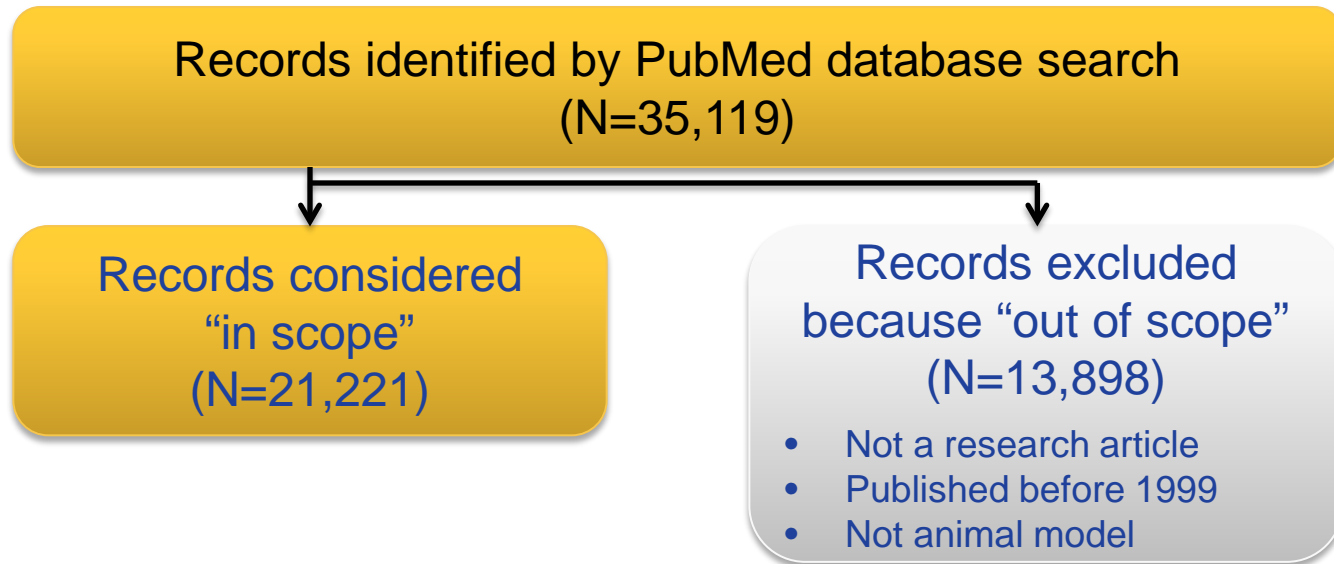
(genome-wide OR genome-wide techniques)

N=35,119

records retrieved



How were records processed?





What topics are in this literature?

	# records
Topic 28: genes, expression, gene, analysis, identified, microarray, expressed, involved, identify, genome-wide	11778
Topic 41: epigenetic, mechanisms, development, gene, genetic, studies, disease, epigenetics, regulation, recent	8637
★ Topic 40: methylation, dna, cpg, methylated, promoter, island, sites, islands, human, regions	7869
Topic 29: cell, cells, growth, signaling, expression, kinase, apoptosis, protein, proliferation, pathway	6793
Topic 11: transcription, promoter, gene, binding, expression, transcriptional, factor, sites, regulation, region	6208
Topic 2: histone, chromatin, epigenetic, modifications, acetylation, gene, modification, methylation, dna, lysine	4933
Topic 21: cancer, prostate, expression, cell, cells, tumor, pancreatic, human, lines, invasion	3491
✕ Topic 38: histone, protein, proteins, complex, domain, mbd, polycomb, ezh, binding, silencing	3392
Topic 48: protein, proteins, mitochondrial, stress, cells, cellular, nuclear, membrane, cell, oxidative	3259

- 50 topics are automatically generated based on the most frequently used words
- Topics were surveyed to get a sense of the types of records that had been retrieved
- Topics were screened and ~120 records were identified to serve as a seed set (60 relevant★ 67 not relevant✕)



How do the records compare to seed sets?

SWIFT File Preferences

SWIFT Text Mining Workbench - [//Users/pelchke/Desktop/SWIFT Release Version 1.0/Data/Global DNA Methylation_3_KP.stp]

Search Browse MeSH Tree Term Browser Document Folders Document Preview Fingerprint Word Cloud Pie Chart Ranking Performance

Document Type Edit...

All Documents

- Review Articles - Ranked
- Reviews
- Not Reviews
- Not Reviews 1999 and later
- has unwanted MeSH
- comp unwanted MeSH
- usable set

Redistribute Prioritize

Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broilers.

Hu Y, Xu H, Li Z, Zheng X, Jia X, Nie Q, Zhang X. *PloS one* (2013)

▼ Abstract

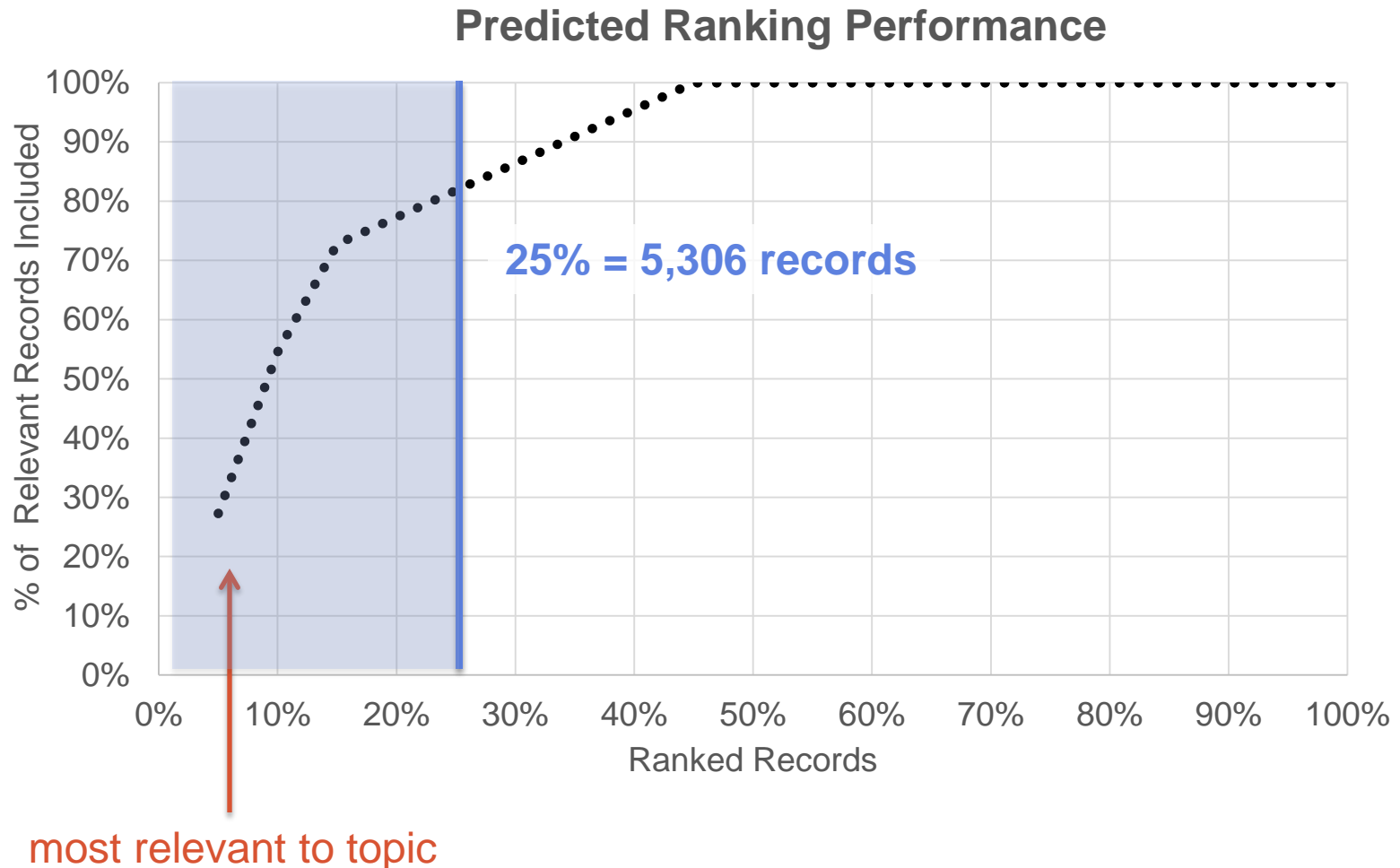
Growth traits are important in poultry production, however, little is known for its regulatory mechanism at epigenetic level. Therefore, in this study, we aim to compare DNA methylation profiles between fast- and slow-growing broilers in order to identify candidate genes for chicken growth. Methylated

Showing 21221 of 35119 loaded documents (60 total included; 127 total training docs.)

Training Item?	Includ...	PMID	Title	Year	Authors
✓	✓	23526956	Integrative genome-wide gene expression profiling of clear cell renal cell carcinoma in Czech Republic and in th...	2013	Wozniak MB, Le Calvez-Keln...
✓	✓	23414147	Genome-wide map of quantified epigenetic changes during in vitro chondrogenic differentiation of primary hu...	2013	Herlofsen SR, Bryne JC, Høib...
✓	✓	24113113	Genome-wide DNA methylation analysis reveals distinct impact of differential methylations on gene transcriptio...	2013	Yu YP, Ding Y, Chen R, Liao...
✓	✓	23411189	Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broilers.	2013	Hu Y, Xu H, Li Z, Zheng X, Ji...
✓	✓	23589626	Laser capture microdissection-reduced representation bisulfite sequencing (LCM-RRBS) maps changes in DNA ...	2013	Schillebeeckx M, Schrade A,
✓	✓	24057373	Global DNA methylation analysis using methyl-sensitive amplification polymorphism (MSAP).	2014	Yaish MW, Peng M, Rothstei...
✓	✗	11524426	Typical demethylation events in cloned pig embryos. Clues on species-specific differences in epigenetic reprog...	2001	Kang YK, Koo DB, Park JS, C...
✓	✗	12461650	The distinguishing sequence characteristics of mouse imprinted genes.	2002	Ke X, Thomas NS, Robinson...
✓	✗	12436242	A novel approach for identifying candidate imprinted genes through sequence analysis of imprinted and control...	2002	Ke X, Thomas NS, Robinson...
✓	✗	12215906	Inactivation of the Fas gene by Alu insertion: retrotransposition in an intron causing splicing variation and autoi...	2002	Tighe PJ, Stevens SE, Demps...
✓	✗	14691535	Identification of Drosophila MicroRNA targets.	2003	Stark A, Brennecke J, Russel...
✓	✗	15480131	Distinct patterns of LINE-1 methylation level in normal tissues and the association with carcinogenesis.	2004	Chalitchagorn K, Shuangsho...
✓	✗	15520080	Analysis of a marsupian MHC region containing two recently duplicated class I loci.	2004	Miska KB, Wright AM, Lunde...
✓	✗	15702417	LINE-1 amplification accompanies explosive genome repatterning in rodents.	2004	Dobigny G, Ozouf-Costaz C...
✓	✗	15969504	MicroRNAs 143 and 145 are possible common onco-microRNAs in human cancers.	2006	Akao Y, Nakagawa Y, Naoe...
✓	✗	17476321	LINE-1 hypomethylation in cancer is highly variable and inversely correlated with microsatellite instability.	2007	Estécio MR, Gharibyan V, Sh...
✓	✗	17823410	MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion...	2007	Corney DC, Flesken-Nikitin...
✓	✗	17603915	Analysis of the features and source gene composition of the AluYg6 subfamily of human retrotransposons.	2007	Styles P, Brookfield JF

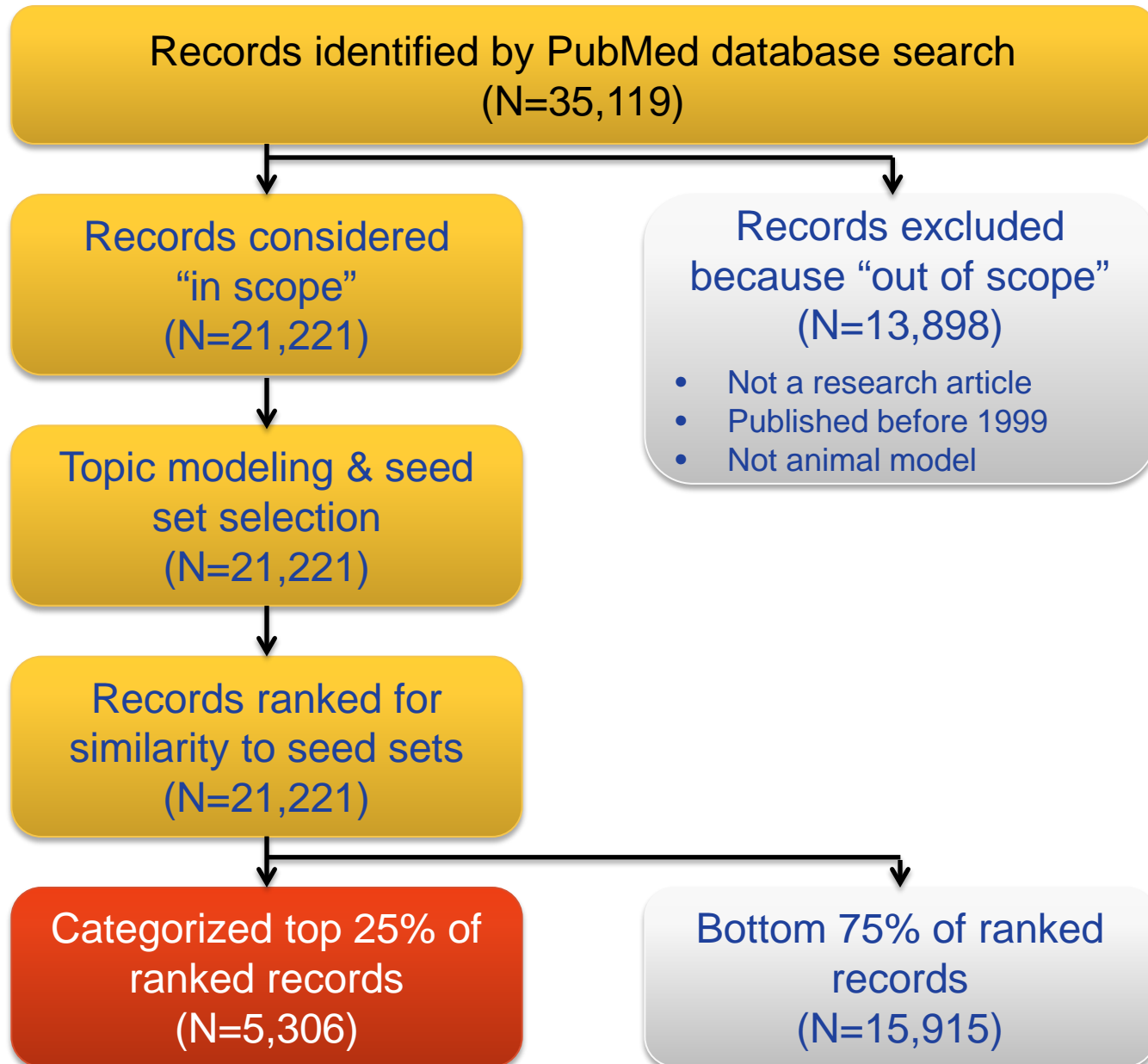


Which were the most relevant records?





How were records processed?



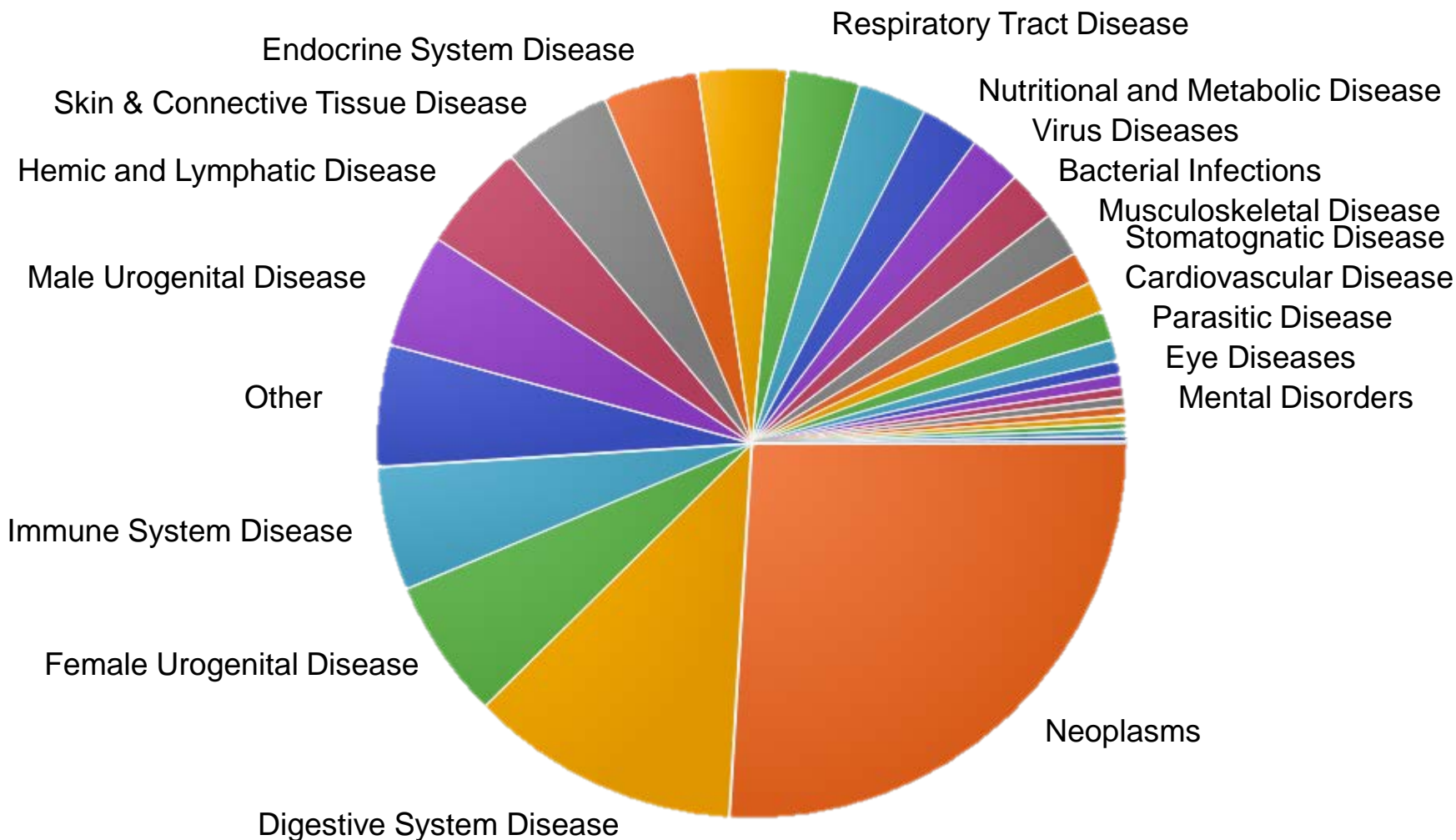


T=title
A=abstract
S=suppl. info
M=MeSH

[illegible]



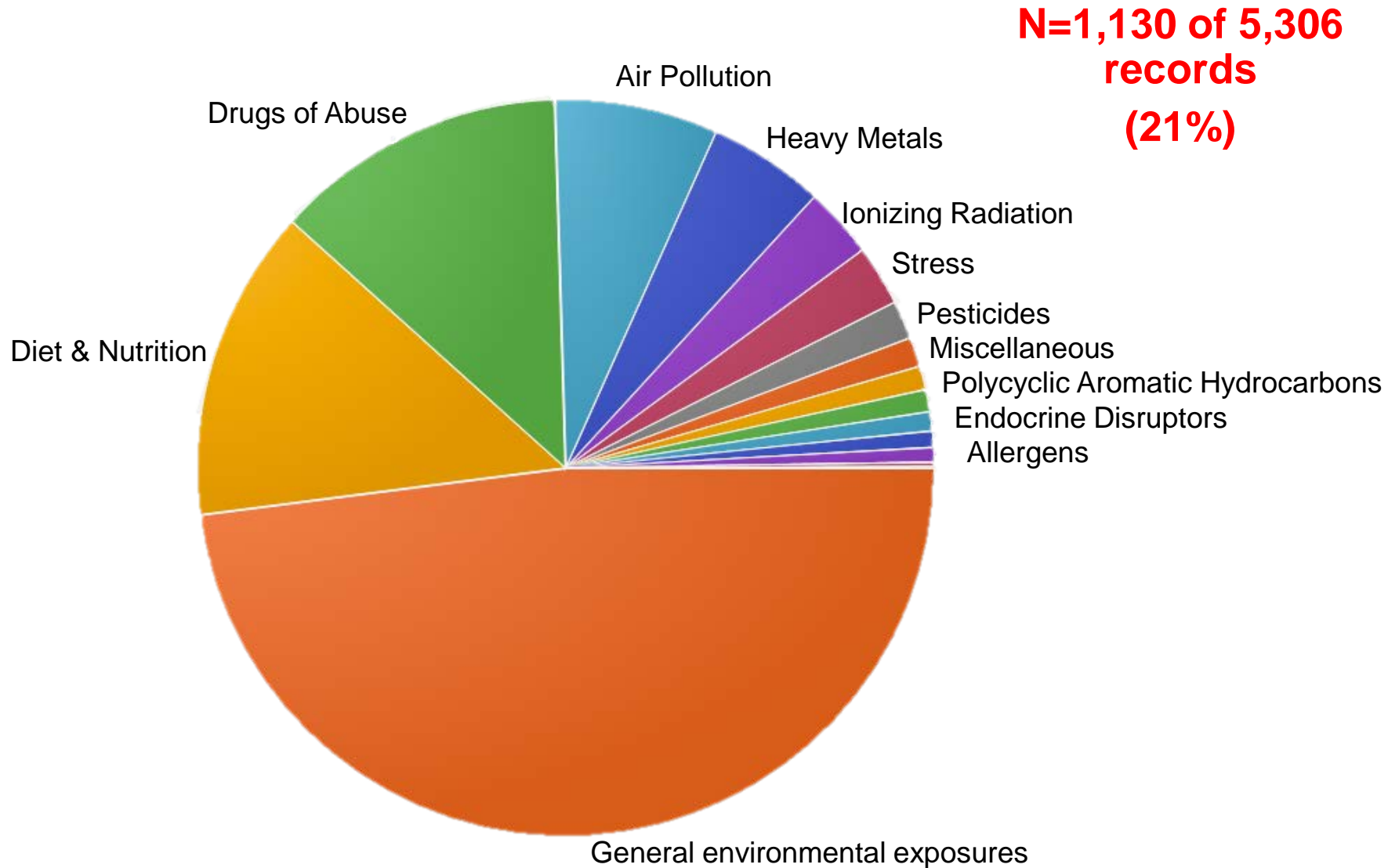
What health outcomes are investigated?



N=5,306 records



What types of exposures are investigated?

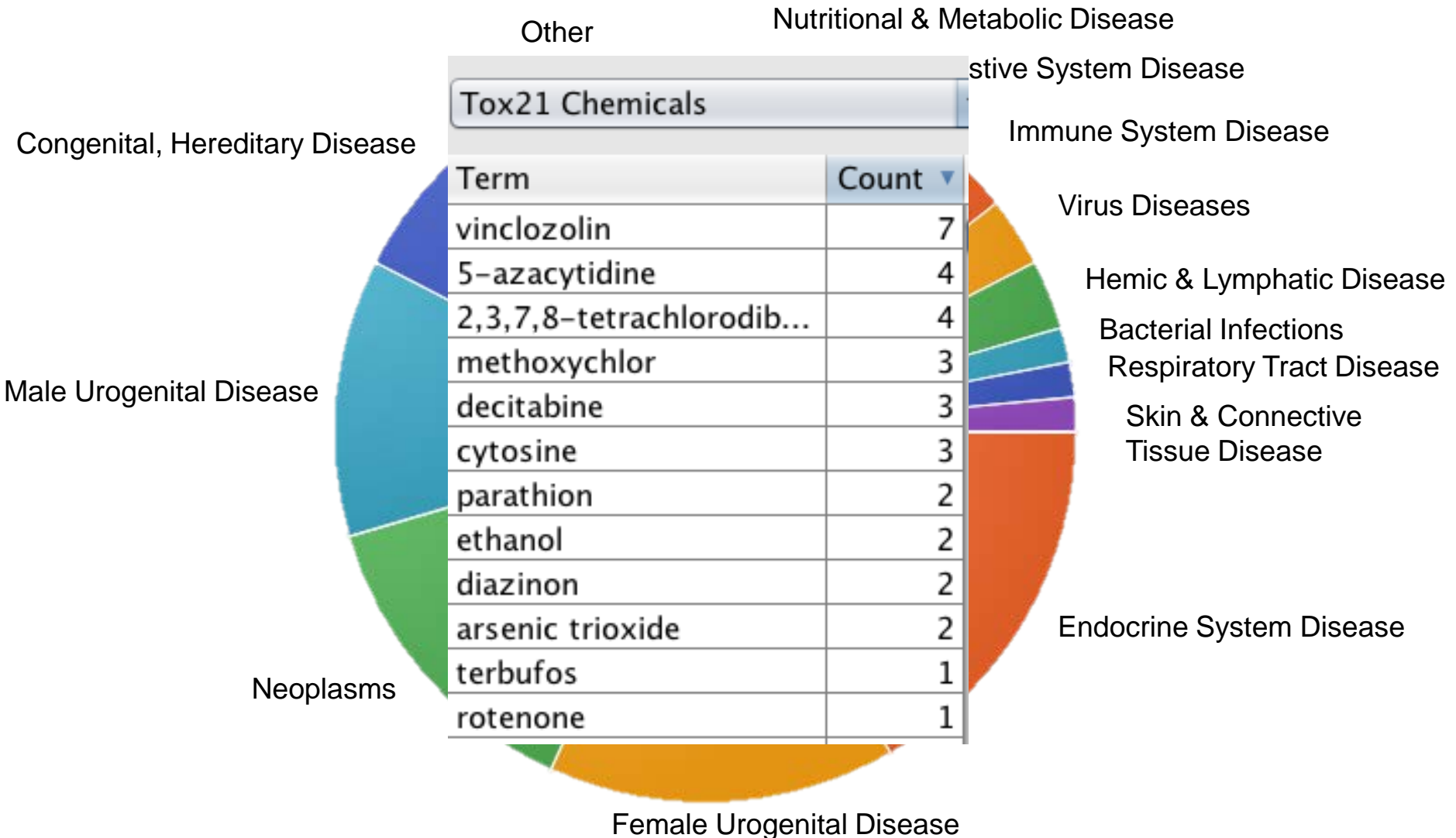




Evaluating exposure & health outcomes

Exposure = pesticide

N=26 records





- Scoping reports useful for large, complex or emerging literatures
 - Get a sense of the literature
 - Highlights the variety in the field
 - Line up future projects
 - Research tool for data mining
- Implemented a new text mining tool
 - Useful for visualizing



Acknowledgements

- **Office of Health Assessment and Translation**

- Abee Boyles
- Kembra Howdeshell
- Katie Pelch
- Andrew Rooney, Deputy Director
- Kyla Taylor
- Kristina Thayer, Director
- Vickie Walker

- **Office of Scientific Information Management**

- Stephanie Holmgren

- **SCIOME, LLC**

- Brian Howard
- Jason Phillips
- Ruchir Shah

- **Evaluation Design Team**

- Douglas Bell
- Lisa Chadwick
- David Fargo
- Stephanie London
- Alex Merrick
- Jack Taylor
- Christina Teng
- Fred Tyson
- Paul Wade
- Lauren Wilson