Published in final edited form as:

Reprod Toxicol. 2011 May ; 31(4): 562–569. doi:10.1016/j.reprotox.2010.11.008.

Drug target-gene signatures that predict teratogenicity are enriched for developmentally related genes

Asher D. Schachter^{1,3,4} and Isaac S. Kohane^{2,3,4}

¹ Division of Nephrology, Children's Hospital Boston, Boston, MA

² Division of Endocrinology, Children's Hospital Boston, Boston, MA

³ Children's Hospital Informatics Program, Boston, MA

⁴ Harvard Medical School, Boston, MA

Abstract

Drugs prescribed during pregnancy affect two populations simultaneously: fetuses and their mothers. Drug-induced fetal injury (teratogenicity) has a significant impact on current and future public health. Teratogenic risk designation of many drugs relies on associating rare fetal events with rare environmental exposures. Therefore we aim to develop preclinical predictive models of clinical teratogenicity. We collated public databases for drug-target-gene relationships for 619 drugs spanning the 5 pregnancy risk classes. Genes targeted by high risk but not low risk drugs demonstrated 79% accuracy ($p < 0.0001$ vs. random) for predicting high vs. low fetal risk on cross validation. Functional enrichment analysis revealed that target genes of drugs known to be safe in pregnancy contained no developmentally related terms, while target genes of known teratogens contained 85 developmentally related terms. Drug target gene signatures that are enriched for known developmental genes may provide valuable preclinical predictive information regarding drug pregnancy risk.

Keywords

teratogen; drug; development; fetus; pregnancy; toxicity; congenital; informatics

1. Introduction

Drug-induced fetal injury (teratogenicity) is a unique and doubly significant public health issue in that it affects two populations, simultaneously: fetuses and their mothers. Pregnant women may be prescribed medications to treat chronic conditions predating conception, such as autoimmune disorders and infection, and may also require medical therapy to manage pregnancy-induced complications such as pre-eclampsia. Obstetrical management decisions must balance several considerations: the effects of the natural course of an untreated underlying disorder on the mother and on the fetus, and the effects of the indicated therapy on the mother and on the fetus. In some instances, treating the disorder benefits both

Corresponding Author: Asher D. Schachter, Division of Nephrology, HU-319, Children's Hospital Boston, 300 Longwood Avenue, Boston, MA 02115, Tel. 617-355-6129, asher.schachter@childrens.harvard.edu.

Conflict of Interest Statement: The authors declare that there are no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

mother and fetus, while in other cases the treatment benefits one and may harm the other. These management decisions are more often hampered by insufficient information regarding a specific medication's effect on the fetus, because formal clinical trials of new drugs in the fetal population are unethical.

Therefore, teratogenicity remains a challenging issue in drug development. The lessons learned from thalidomide use in the 1960s [1,2] still resonate today. Currently, teratogenicity categorization consists of 5 classes: A (safest), B (little evidence of fetal risk), C (animal evidence of fetal risk and/or human data unavailable), D (evidence of human risk but benefits may outweigh risks in some cases), and X (known teratogen). However, there is insufficient preclinical and clinical evidence to classify the vast majority of drugs accurately [3]. Historically, the designation of many drugs as teratogens has relied on “the astute clinician” method [4], in which the association of a rare fetal event with a rare environmental exposure identifies a given substance as a teratogen. The inability to detect more common teratogens, and the need for some infants to be affected for detection limits this approach. It is likely that the majority of teratogenicity occurs unrecognized due to more subtle drug-induced malformations [5]. Clinical testing of new medications for teratogenicity in pregnant women is clearly not an option. As such, high-throughput innovations of previously described “predictive teratology” approaches [6,7] would be of tremendous value in assessing developmental toxicity risks. The identification of relevant mutations and polymorphisms is central to the innovation of pre-pregnancy teratogen risk assays [8].

Yao and Rzhetsky recently reported results of a network analysis they performed on drug targets for 919 drugs [9]. Among their findings was the observation that drug targets that are also essential genes are more likely to injure embryonic tissue. Given the likelihood that more common teratogenic effects may manifest as subtle anomalies [5], we posited that developmentally related functions could be over-represented in the groups of (mostly) non-essential genes within target protein families (PFAM(s)) of teratogenic drugs.

We have collated and integrated several public drug and genomic databases for drug-target PFAM relationships, and PFAM-gene relationships encompassing 619 drugs spanning the 5 pregnancy risk classes. In addition, given the lack of evidence supporting the current pregnancy risk classification scheme, we performed a similar analysis on 3670 drugs irrespective of known pregnancy risk class. Our objective was to determine whether drug-target-gene data would inform pregnancy risk class. This approach may assist in determining a new drug's teratogenicity risk prior to first-in-human trials, since drug targets can be determined in the preclinical phase of drug development.

2. Materials And Methods

2.1 Data

We downloaded all “drug cards” from DrugBank [10,11], encompassing data for 4774 drugs, and reformatted the data for analysis in the R programming environment [12]. The DrugBank data includes the PFAM identification numbers (IDs) for most drugs' target(s). We downloaded pregnancy risk classification for unique compounds in the www.safefetus.com website matching the 4774 drugbank.ca drugs, yielding a total 720 common drugs. As some drugs are designated by more than one risk class depending on the clinical situation, we assigned numeric values (1,2,3,4,5) to each class, (A, B, C, D, X, respectively), and took the “mean class”, rounding up to the higher risk class when a fraction resulted. We accessed Bioconductor's org.Hs.egPFAM annotation database [13] to determine the GenBank gene IDs for the genes belonging to each PFAM.

2.2 Cross Validation

We performed cross-validation on randomly selected subsets of the 14 Class A and 45 Class X drugs. For each of 10 iterations, we used using R's *sample()* function to randomly select 2 Class A and 4 Class X drugs to serve as the test set, leaving a training set of 12 Class A and 41 Class X drugs. Within each iteration we selected the genes belonging to the target PFAMs for the training set Class X drugs ("Class X training genes"). For each test set drug we then calculated the Gene Risk Quotient (GRQ), defined as the quotient of the number of a drug's Class-X-not-A genes over the drug's total number of target PFAM genes. To calculate the GRQ we used the number of "Class X training genes" present in the test drug's target gene list, and the total number of genes present in the test drug's target gene list. We repeated each set of 10 iterations 50 times in order to calculate mean and standard deviation values for the performance metrics. We used using R's *sample()* function to randomly select 1023 genes from the training set drugs in each iteration to serve as a parallel, random control group.

We calculated receiver operator characteristic (ROC) curve coordinates using the predicted GRQ values, which are continuous numbers between 0 and 1. We used the c-index method [14] to calculate the area under each iteration's ROC curve. Briefly, this formula first requires counting the number of concordant pairs and the number of ties across all possible pairs of predicted output values. The c-index is then calculated by the formula (number concordant + 0.5 * number of ties) / total number of pairs.

2.3 Functional Enrichment Analysis

For functional enrichment analysis, we used R's *hyperGTest()* function (*GOstats* package), which computes hypergeometric p-values for over or under-representation of each term in GO categories among the specified gene set. A *GOHyperGParams* class object was created for each of the two extreme Classes (A and X), setting the *GO ontology* parameter to "BP" (for biologic process), the *universeGeneIds* parameter to include all 7426 unique genes, and the *annotation* parameter to "org.Hs.eg.db". For the Class A *GOHyperGParams* object we set the *geneIds* parameter to include genes present in class A and not X (1114 genes). For the Class X *GOHyperGParams* object we set the *geneIds* parameter to the 1023 Class-X-not-A genes. We used R's *agrep()* function to perform fuzzy matching of GO terms to developmentally related words ("genesis", "develop", "differentiat", or "growth").

2.4 Global Drug Analysis

Given that current pregnancy risk categorization is inaccurate and often estimated based on little or no data, we performed functional enrichment analyses to assess for over-representation of developmentally related GO terms in all target PFAM genes for all drugs with known PFAM in the DrugBank database. For each drug, we determined the target(s) PFAM ID(s), and all gene IDs in the drug's target's PFAM were extracted from Bioconductor's org.Hs.egPFAM annotation database. Then each drug's target PFAM genes were submitted to R's *hyperGTest()* function (*GOstats* package). We performed fuzzy matching of GO terms to developmentally related words, as described above.

3. Results

3.1 The proportion of drugs targeting Class-X-not-A genes increases in parallel with the teratogenic risk

Of the 4774 drugs that we downloaded from DrugBank [10,11], pregnancy risk classification as per the SafeFetus website (www.safefetus.com) was available for a total 720 common drugs. Of the 720 drugs, 619 had known target protein family (PFAM) IDs, including 14, 139, 325, 96, and 45 drugs in Classes A, B, C, D, and X, respectively.

Collectively, the 619 drugs were associated with 578 unique PFAM IDs. We accessed Bioconductor's org.Hs.egPFAM annotation database [13] to obtain 7426 unique GenBank gene IDs for the gene products with domains belonging to each of the 578 PFAMs. Of the 7426 genes, 1114 genes were targeted by class A but not class X drugs, and 1023 genes were targeted by class X but not class A drugs. Figure 1A shows the distribution of the 1023 drugs targeting Class-X-not-A genes across the 5 teratogen risk classes. Figure 1B summarizes the “gene risk quotient” (GRQ) values (defined as the quotient of the number of a drug's Class-X-not-A genes over the drug's total number of target PFAM genes) for each drug, stratified by teratogen risk class. Both figures show that the proportion of drugs targeting class-X-not-A target genes increases in parallel with current teratogenic risk designation. The Pearson correlation coefficient for the relationship of teratogenicity risk class vs. proportion of drugs targeting ≥ 1 class X-not-A gene (Figure 1A) is 0.92 ($p=0.02$). The Pearson correlation coefficient for the relationship of teratogenicity risk class vs. GRQ (Figure 1B) is 0.39 ($p=0.05$).

3.2 Preclinical predictive modeling of drug-gene dependencies and known teratogenic risk

The drug-gene dependencies appear to have significant predictive accuracy regarding teratogenicity. Cross validation results are displayed in Figure 2, which shows the mean and 95% confidence interval receiver-operator characteristic (ROC) curves for the cross-validation of the drugs targeting Class-X-not-A genes (Figure 2A), yielding an area under the curve (AUC) of 0.8 ± 0.06 (mean $\pm$ SD), and accuracy of 0.79 ± 0.07 . In contrast, the randomly selected genes performed with an ROC AUC of 0.45 ± 0.06 and accuracy of 0.35 ± 0.06 (Figure 2B). The two ROC curves are statistically significantly different ($p < 0.0001$, paired t-test).

3.3 Functional enrichment analysis of drug target gene signatures

We performed functional enrichment analyses in order to determine whether each drug's PFAM target genes represented specific biologic processes. Hypergeometric analysis yielded 196 unique Gene Ontology (GO) terms for Class A-not-X genes (p -value range $3.6 \times 10^{-64} - 9.9 \times 10^{-3}$), and 493 unique GO terms for Class X-not-A genes (p -value range $1.0 \times 10^{-56} - 9.9 \times 10^{-3}$). Fuzzy matching of GO terms to any of “genesis”, “develop”, “differentiat”, or “growth” yielded no GO terms matching for Class A-not-X, and 85 GO terms matching for Class X-not-A (Table 1). The 140 Class X-not-A genes represented by these 85 GO terms shown in Table 2 include 18 collagen genes, 9 bone morphogenetic protein genes, 8 tumor necrosis family genes, 8 growth differentiation factor genes, 8 SLIT/slit homolog genes, 6 RAS oncogenes, 5 kallikrein-related peptidase genes, 4 matrix metalloproteinase genes, 3 c-fos/vascular endothelial growth factor genes, 3 inhibin genes, and 3 transforming growth factor beta genes. Table 3 summarizes the over-represented developmental GO terms that are unique to each class. These are the developmental GO terms associated only with the specified pregnancy risk class.

Given that current pregnancy risk categorization is inaccurate and is often estimated based on a paucity of data, we performed functional enrichment analyses to assess for over-representation of developmentally related GO terms in all target PFAM genes for all drugs in the DrugBank database. Of the 4774 DrugBank drugs, 4316 drugs were unique (non-duplicate). For 3670 of 4316 drugs, target PFAM genes had corresponding GO terms for the biologic process ontology. For each of the 3670 drugs, the ratio of the number of developmentally related GO terms over the total number of terms was calculated. 1881 of these 3670 drugs (51%) had non-zero ratios (at least one developmentally-related GO term). Across these 1881 drugs, the ratio of developmental GO terms to total GO terms had a mean $\pm$ standard deviation of 0.114 and 0.109, respectively, and a median and interquartile range of 0.0681 and 0.0197 – 0.195, respectively. Figure 3 shows these ratios stratified by

pregnancy risk class, for the 1881 drugs for which the ratio was greater than zero. Of these 1881 drugs with non-zero ratios, drugs with unknown pregnancy risk class comprise the largest single group (1325 drugs, 70%).

4. Discussion

The clinical pregnancy risk classification of many drugs is a challenge [3], relying on astute clinicians to detect and report associations between rare fetal phenotypes and rare environmental exposures [4]. The limitations of the astute clinician approach become even more apparent given the likelihood that rare phenotypes represent a small proportion of all fetal drug-induced malformations [5]. It is extremely challenging to assess clinical teratogenicity of drugs, particularly in any prospective manner. Nevertheless, with the ongoing advances in high-risk pregnancy management, increasingly larger numbers of pregnant women will be requiring and receiving prescription medications. For many of these medications, the effects on the fetus are currently unknown and may also be relatively subtle, further complicating clinical identification. Our study is aimed at determining gene signatures that could potentially be used to flag drugs that target these gene products.

In Figure 1, we show that the proportion of drugs targeting Class-X-not-A genes increases in parallel with current teratogenic risk classification. The Gene Risk Quotient (GRQ) was trained on Class A and Class X drugs, and therefore the difference in GRQ between these two groups is expected. However, these results are intriguing because the GRQ is higher in Class D (likely unsafe) vs. Class B (likely safe) drugs. In Class C in which clinical (human) teratogenicity remains unknown it is therefore possible that the GRQ will target gene signatures to potentially provide an initial filter to flag Class C drugs that may require additional scrutiny in terms of additional preclinical teratogenicity studies and post-approval surveillance for more subtle teratogenic effects. Furthermore, since the current pregnancy risk classification scheme is considered unreliable, particularly for the majority of drugs that comprise classes B, C, and D, we performed an “unsupervised” analysis, i.e. independent of teratogen risk class. We have shown that of the 1881 drugs that target genes functionally related to at least one developmental GO term, 70% (1325 drugs) are of unknown teratogen risk class, and an additional 355 drugs (19%) fall into Class C risk (human data unavailable). Taken together, this represents 1680 (89%) drugs that (a) target developmental genes, and (b) have unknown effects on the human fetus. The preclinical target gene signature and GRQ may therefore enable more efficient preclinical teratogenicity screening for a large number of compounds, and might serve as another tool to help focus limited resources for post-approval surveillance of teratogenicity.

In order to determine the value of the GRQ for predicting known teratogen risk (Class A vs. Class X only), in Figure 2 we show that drug target gene signatures provide predictive accuracy of 79% for known teratogenic risk on cross validation (see Materials and Methods), and in comparison to random gene signatures. It is therefore possible that the GRQ may predict Class C teratogenicity with a similar degree of accuracy. These results further underscore the potential value of the GRQ, as discussed above.

Predictive markers may provide accurate information independent of functional relationships, and therefore we performed functional enrichment analysis to assess functional relevance for the gene signatures targeted by specific drug risk classes. Our finding that GO terms from Class A-not-X target PFAM genes contained no developmentally related terms while Class X-not-A target PFAM genes contained 85 developmentally related GO terms supports the hypothesis that drug-induced fetal malformations are may be mediated by known developmental genes. This is an intriguing observation because it is logical to hypothesize that a proportion of teratogenic medications

disrupt fetal development by altering developmentally implicated gene products, and that drugs that are safe in pregnancy do not target developmentally sensitive genes. Furthermore, specific classes may actually approximate clusters of drugs that target developmentally specific systems. As Table 3 shows, fundamental growth, development, and embryogenesis processes are over-represented in Class X along with gastrointestinal processes, whereas other classes also appear to affect predominantly specific organ systems in addition to fundamental developmental processes: central nervous system (Class B); musculoskeletal, immune, and cutaneous systems (Class C); and cardiovascular and immune systems (Class D). However, these results should be interpreted with caution given the limitations of the current teratogenicity classification system. For example, a drug currently classified in Class C (human data unavailable) may be reclassified to a different class in the future when more human data becomes available.

Figure 3 shows that median G.O. developmental term ratios are similar across classes B, C, D, X, and drugs that are not classified (N/A). There are several possible explanations for this. The current pregnancy risk classification system is unreliable, as we discussed in Section 1. Introduction. Another limitation of this system is that some drugs are listed in more than one class, depending on the specific clinical situation. This presents somewhat of a circular problem for our approach in that we cannot reliably verify class predictions for the vast majority of drugs. An additional limitation of our approach is the inability to assess as yet unknown target pathways: our approach assumes that each drug's known targets encompass all of the drug's targets. Furthermore, we are unable to determine any effects of dosing on teratogenicity. At lower therapeutic doses some drugs may be safe or may induce milder and potentially undetectable teratogenic effects, while at higher doses within the therapeutic range these same drugs may be teratogenic. Transport across the placental barrier may also be affected by dose. Drug interactions are also not assessed by our approach. Some drugs may be safe when administered alone, but may reach teratogenic levels in the maternal bloodstream if administered in conjunction with a second medication that inhibits the first drug's metabolic breakdown. Maternal co-morbidities such as pre-eclampsia, diabetes, and renal insufficiency may also affect a medication's pharmacokinetics and teratogenic potential.

In summary, we used public resources to identify a unique, teratogen-specific, developmentally rich gene signature of genes. Drug target gene signatures that are enriched for known developmental genes may provide valuable preclinical predictive information regarding clinical drug pregnancy risk. It may therefore be possible to identify potential teratogens and conduct additional animal studies prior to first-in-human trials, and to focus post-approval surveillance efforts on higher risk drugs, with significant public health benefits to pregnant women and their fetuses.

Acknowledgments

The authors are grateful for the advice provided by Dr. Lewis B. Holmes. This work was supported by NIH grant R01GM085421 (ADS). Author contributions: ADS conceived the idea for this work, designed the analysis, acquired & formatted the data, performed the analyses, and wrote the manuscript. ISK assisted in the design of the study, and proofread all drafts of the manuscript. The authors dedicate this work to our friend and colleague, Marco Francesco Ramoni, PhD, who inspired so many with his dedication, insights, teachings and friendship.

References

1. Mellin GW, Katzenstein M. The saga of thalidomide. Neuropathy to embryopathy, with case reports of congenital anomalies. *N Engl J Med*. 1962; 267:1238–44. concl. [PubMed: 13934700]
2. Stempel E. The impetus of thalidomide on drug legislation and regulation. *Am J Pharm Sci Support Public Health*. 1962; 134:355–64. [PubMed: 13983817]

3. Lo WY, Friedman JM. Teratogenicity of recently introduced medications in human pregnancy. *Obstet Gynecol.* 2002; 100:465–73. [PubMed: 12220765]
4. Carey JC, Martinez L, Balken E, Leen-Mitchell M, Robertson J. Determination of human teratogenicity by the astute clinician method: review of illustrative agents and a proposal of guidelines. *Birth Defects Res A Clin Mol Teratol.* 2009; 85:63–8. [PubMed: 19107954]
5. Holmes L, Harvey E, Coull B, Huntington K, Khoshbin S, Hayes A, et al. The teratogenicity of anticonvulsant drugs. *N Engl J Med.* 2001; 344:1132–8. [PubMed: 11297704]
6. Fabro S, Shull G, Brown NA. The relative teratogenic index and teratogenic potency: proposed components of developmental toxicity risk assessment. *Teratog Carcinog Mutagen.* 1982; 2:61–76. [PubMed: 6122280]
7. Augustine-Rauch KA. Predictive teratology: teratogenic risk-hazard identification partnered in the discovery process. *Curr Drug Metab.* 2008; 9:971–7. [PubMed: 18991594]
8. Van Dyke DC, Ellingrod VL, Berg MJ, Niebyl JR, Sherbondy AL, Trembath DG. Pharmacogenetic screening for susceptibility to fetal malformations in women. *Ann Pharmacother.* 2000; 34:639–45. [PubMed: 10852093]
9. Yao L, Rzhetsky A. Quantitative systems-level determinants of human genes targeted by successful drugs. *Genome Res.* 2008; 18:206–13. [PubMed: 18083776]
10. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006; 34:D668–72. [PubMed: 16381955]
11. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008; 36:D901–D6. [PubMed: 18048412]
12. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2009.
13. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004; 5:R80. [PubMed: 15461798]
14. Tsuruta H, Bax L. Polychotomization of continuous variables in regression models based on the overall C index. *BMC Med Inform Decis Mak.* 2006; 6:1–13. [PubMed: 16396679]

Abbreviations

PFAM IDs	protein family identification numbers
GO	gene ontology
GRQ	gene risk quotient
ROC AUC	receiver operator characteristic, area under the curve

Fig. 1. Distribution of the 1023 Class-X-not-A genes (A) and Gene Risk Quotient values (B) across the 5 teratogen risk classes. **A:** Bargraph showing the proportion of the drugs in each class that target one or more Class-X-not-A genes. Note that the proportion increases in parallel with the teratogenic risk (Pearson correlation coefficient = 0.92, $p = 0.02$). **B:** Boxplot showing the medians and interquartile ranges for the "gene risk quotient" (GRQ) values for each drug, stratified by teratogen risk class (Pearson correlation coefficient = 0.39, $p = 0.05$). The GRQ is defined as the quotient of the number of a drug's Class-X-not-A genes over the drug's total number of targeted genes.

Fig. 2.

Receiver operator characteristic (ROC) curves for cross validation of (A) the Class-X-not-A gene Gene Risk Quotient (GRQ), and (B) the GRQ for randomly sampled parallel genes sets. The Class-X-not-A gene GRQ values provide accurate prediction of drug teratogenicity in comparison to randomly selected genes/AUC= area under the curve. Values are mean $\pm$ standard deviation. The (paired) student t-test was used for p-value calculation.

Fig. 3.

Boxplot of the ratios of the number of developmentally related GO terms over the total number of terms for the 1881 of 3670 drugs (51%) that had non-zero ratios (at least one developmentally-related GO term). The ratios are stratified by pregnancy risk class. Drugs with unknown pregnancy risk class comprise the largest group, with 1325 drugs.

Table 1

Developmentally related Gene Ontology terms of the genes targeted by Class X but not Class A drugs

GO ID	Term
GO:0009653	anatomical structure morphogenesis
GO:0022603	regulation of anatomical structure morphogenesis
GO:0009887	organ morphogenesis
GO:0032989	cellular component morphogenesis
GO:0042476	odontogenesis
GO:0000902	cell morphogenesis
GO:0022604	regulation of cell morphogenesis
GO:0000904	cell morphogenesis involved in differentiation
GO:0007435	salivary gland morphogenesis
GO:0010769	regulation of cell morphogenesis involved in differentiation
GO:0022008	neurogenesis
GO:0022612	gland morphogenesis
GO:0050767	regulation of neurogenesis
GO:0007409	axonogenesis
GO:0048667	cell morphogenesis involved in neuron differentiation
GO:0048812	neuron projection morphogenesis
GO:0048858	cell projection morphogenesis
GO:0010770	positive regulation of cell morphogenesis involved in differentiation
GO:0032990	cell part morphogenesis
GO:0045765	regulation of angiogenesis
GO:0001525	angiogenesis
GO:0048646	anatomical structure formation involved in morphogenesis
GO:0048514	blood vessel morphogenesis
GO:0016525	negative regulation of angiogenesis
GO:0042481	regulation of odontogenesis
GO:0050793	regulation of developmental process
GO:0048856	anatomical structure development
GO:0048731	system development
GO:0048513	organ development

GO ID	Term
GO:0009888	tissue development
GO:0001501	skeletal system development
GO:0051093	negative regulation of developmental process
GO:0032502	developmental process
GO:0048869	cellular developmental process
GO:0051216	cartilage development
GO:0051094	positive regulation of developmental process
GO:0043588	skin development
GO:0007275	multicellular organismal development
GO:0060284	regulation of cell development
GO:0048666	neuron development
GO:0035272	exocrine system development
GO:0031175	neuron projection development
GO:0060348	bone development
GO:0008544	epidermis development
GO:0048468	cell development
GO:0007431	salivary gland development
GO:0051960	regulation of nervous system development
GO:0007398	ectoderm development
GO:0014031	mesenchymal cell development
GO:0060485	mesenchyme development
GO:0048535	lymph node development
GO:0051797	regulation of hair follicle development
GO:0050773	regulation of dendrite development
GO:0001568	blood vessel development
GO:0001944	vasculature development
GO:0010720	positive regulation of cell development
GO:0010975	regulation of neuron projection development
GO:0002520	immune system development
GO:0007517	muscle organ development
GO:0001541	ovarian follicle development

GO ID	Term
GO:0048589	developmental growth
GO:0051798	positive regulation of hair follicle development
GO:0045595	regulation of cell differentiation
GO:0002573	myeloid leukocyte differentiation
GO:0002761	regulation of myeloid leukocyte differentiation
GO:0030225	macrophage differentiation
GO:0030154	cell differentiation
GO:0045637	regulation of myeloid cell differentiation
GO:0002762	negative regulation of myeloid leukocyte differentiation
GO:0045650	negative regulation of macrophage differentiation
GO:0045638	negative regulation of myeloid cell differentiation
GO:0045649	regulation of macrophage differentiation
GO:0045596	negative regulation of cell differentiation
GO:0030099	myeloid cell differentiation
GO:0002521	leukocyte differentiation
GO:0030182	neuron differentiation
GO:0048762	mesenchymal cell differentiation
GO:0010742	foam cell differentiation
GO:0010743	regulation of foam cell differentiation
GO:0010745	negative regulation of foam cell differentiation
GO:0045670	regulation of osteoclast differentiation
GO:0030851	granulocyte differentiation
GO:0007179	transforming growth factor beta receptor signaling pathway
GO:0040007	growth
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway

Table 2

Names for the 140 developmentally related genes in the Class-X-not-A gene group, and the Gene Ontology terms to which each gene mapped.

Gene Name	Mapped Gene Ontology Term
adhesion molecule with Ig-like domain 1	multicellular organismal development, cell differentiation
adiponectin, C1Q and collagen domain containing	negative regulation of foam cell differentiation
anti-Mullerian hormone	cell differentiation
arylsulfatase E (chondrodysplasia punctata 1)	skeletal system development
ATP-binding cassette, sub-family A (ABC1), member 1	negative regulation of foam cell differentiation
ATP-binding cassette, sub-family G (WHITE), member 1	negative regulation of foam cell differentiation
bone gamma-carboxyglutamate (gla) protein	odontogenesis, skeletal system development, regulation of osteoclast differentiation
bone morphogenetic protein 10	multicellular organismal development, growth
bone morphogenetic protein 2	organ morphogenesis, cartilage development, growth
bone morphogenetic protein 3	cartilage development, multicellular organismal development, cell differentiation
bone morphogenetic protein 4	angiogenesis, cartilage development, growth
bone morphogenetic protein 5	cartilage development, cell differentiation, growth
bone morphogenetic protein 6	cartilage development, multicellular organismal development, growth
bone morphogenetic protein 7	organ morphogenesis, salivary gland morphogenesis, anatomical structure formation involved in morphogenesis, cartilage development, cell development, growth
bone morphogenetic protein 8a	cartilage development, multicellular organismal development, cell differentiation, growth
bone morphogenetic protein 8b	cartilage development, multicellular organismal development, cell differentiation, growth
caspase 14, apoptosis-related cysteine peptidase	epidermis development, cell differentiation
cell division cycle 42 (GTP binding protein, 25kDa)	macrophage differentiation
cellular retinoic acid binding protein 1	multicellular organismal development
cellular retinoic acid binding protein 2	epidermis development
c-fos induced growth factor (vascular endothelial growth factor D)	angiogenesis, multicellular organismal development, cell differentiation
chemokine (C-C motif) ligand 17	multicellular organismal development
chemokine (C-C motif) ligand 2	organ morphogenesis, transforming growth factor beta receptor signaling pathway
chemokine (C-X-C motif) ligand 10	muscle organ development
chemokine (C-X-C motif) ligand 13	lymph node development
coagulation factor II (thrombin)	multicellular organismal development

Gene Name	Mapped Gene Ontology Term
collagen, type I, alpha 1	blood vessel development
collagen, type I, alpha 2	odontogenesis, skeletal system development, blood vessel development, transforming growth factor beta receptor signaling pathway
collagen, type II, alpha 1	cartilage development
collagen, type III, alpha 1	skin development, transforming growth factor beta receptor signaling pathway
collagen, type IV, alpha 2	negative regulation of angiogenesis
collagen, type IV, alpha 3 (Goodpasture antigen)	negative regulation of angiogenesis
collagen, type IX, alpha 1	organ morphogenesis
collagen, type IX, alpha 2	skeletal system development
collagen, type V, alpha 1	skin development
collagen, type V, alpha 2	skin development
collagen, type V, alpha 3	skin development
collagen, type X, alpha 1	skeletal system development
collagen, type XI, alpha 2	cartilage development
collagen, type XIII, alpha 1	multicellular organismal development, cell differentiation
collagen, type XIX, alpha 1	skeletal system development, multicellular organismal development, muscle organ development, cell differentiation
collagen, type XV, alpha 1	angiogenesis, multicellular organismal development, cell differentiation
collagen, type XVII, alpha 1	epidermis development
collagen, type XVIII, alpha 1	organ morphogenesis
colony stimulating factor 3 (granulocyte)	multicellular organismal development, granulocyte differentiation
complement component 1, q subcomponent, C chain	negative regulation of macrophage differentiation
corin, serine peptidase	anatomical structure morphogenesis
C-reactive protein, pentraxin-related	negative regulation of foam cell differentiation
decorin	organ morphogenesis
ectodysplasin A	multicellular organismal development, ectoderm development, cell differentiation
ectodysplasin A2 receptor	epidermis development, cell differentiation
Fas (TNF receptor superfamily, member 6)	regulation of myeloid cell differentiation
fatty acid binding protein 1, liver	organ morphogenesis
fatty acid binding protein 5 (psoriasis-associated)	epidermis development

Gene Name	Mapped Gene Ontology Term
flightless I homolog (Drosophila)	multicellular organismal development
folliculin	multicellular organismal development, positive regulation of hair follicle development
glial cell derived neurotrophic factor	neuron projection development
ghrelin	multicellular organismal development, cell differentiation
glycoprotein Ib (platelet), alpha polypeptide	cell morphogenesis
growth differentiation factor 10	skeletal system development, transforming growth factor beta receptor signaling pathway
growth differentiation factor 11	organ morphogenesis, skeletal system development, negative regulation of cell differentiation, growth
growth differentiation factor 15	transforming growth factor beta receptor signaling pathway
growth differentiation factor 2	growth
growth differentiation factor 5	transforming growth factor beta receptor signaling pathway, growth
growth differentiation factor 6	growth
growth differentiation factor 7	growth
growth differentiation factor 9	transforming growth factor beta receptor signaling pathway
hepatocyte growth factor (hepatopoietin A; scatter factor)	cell morphogenesis
HtrA serine peptidase 2	neuron development
inhibin, alpha	skeletal system development, ovarian follicle development, cell differentiation, negative regulation of macrophage differentiation
inhibin, beta A	skeletal system development, ovarian follicle development, cell differentiation, negative regulation of macrophage differentiation, growth
inhibin, beta B	ovarian follicle development, cell differentiation, growth
integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	negative regulation of cell differentiation
integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	negative regulation of foam cell differentiation
interleukin 6 (interferon, beta 2)	regulation of angiogenesis, neuron projection development
interleukin 8	angiogenesis
kallikrein-related peptidase 3	negative regulation of angiogenesis
kallikrein-related peptidase 5	epidermis development
kallikrein-related peptidase 6	regulation of cell differentiation
kallikrein-related peptidase 7	epidermis development
kallikrein-related peptidase 8	neuron projection morphogenesis
left-right determination factor 1	multicellular organismal development, transforming growth factor beta receptor signaling pathway

Gene Name	Mapped Gene Ontology Term
left-right determination factor 2	multicellular organismal development, transforming growth factor beta receptor signaling pathway
lymphotoxin alpha (TNF superfamily, member 1)	lymph node development
lymphotoxin beta (TNF superfamily, member 3)	skin development, lymph node development
matrix Gla protein	multicellular organismal development, cell differentiation
matrix metalloproteinase 11 (stromelysin 3)	multicellular organismal development
matrix metalloproteinase 14 (membrane-inserted)	angiogenesis, ovarian follicle development
matrix metalloproteinase 19	angiogenesis, multicellular organismal development, cell differentiation
matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	skeletal system development, macrophage differentiation
muscle RAS oncogene homolog	multicellular organismal development, muscle organ development
myostatin	muscle organ development, transforming growth factor beta receptor signaling pathway, growth
nerve growth factor receptor (TNFR superfamily, member 16)	skin development, multicellular organismal development, cell differentiation
neurotrophic tyrosine kinase, receptor, type 2	multicellular organismal development, regulation of dendrite development
neurotrophic tyrosine kinase, receptor, type 3	multicellular organismal development, cell differentiation
neurturin	neuron projection development
nodal homolog (mouse)	anatomical structure formation involved in morphogenesis, multicellular organismal development
placental growth factor	angiogenesis, multicellular organismal development, cell differentiation
plasminogen	negative regulation of angiogenesis
platelet factor 4	negative regulation of angiogenesis
platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)	transforming growth factor beta receptor signaling pathway
progesterone-associated endometrial protein	multicellular organismal development
proline/arginine-rich end leucine-rich repeat protein	skeletal system development
RAB26, member RAS oncogene family	exocrine system development
RAB3A, member RAS oncogene family	axonogenesis
ras homolog gene family, member A	cell morphogenesis, cell differentiation
ras homolog gene family, member B	multicellular organismal development, cell differentiation
ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	anatomical structure morphogenesis
ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	neuron projection development
receptor tyrosine kinase-like orphan receptor 2	multicellular organismal development
reticulon 4 receptor	axonogenesis

Gene Name	Mapped Gene Ontology Term
retinol binding protein 2, cellular	epidermis development
ribonuclease/angiogenin inhibitor 1	regulation of angiogenesis
RNA binding motif protein 45	multicellular organismal development, cell differentiation
serine peptidase inhibitor, Kazal type 5	negative regulation of angiogenesis
SLIT and NTRK-like family, member 1	axonogenesis
SLIT and NTRK-like family, member 2	axonogenesis
SLIT and NTRK-like family, member 3	axonogenesis
SLIT and NTRK-like family, member 4	axonogenesis
SLIT and NTRK-like family, member 5	axonogenesis
SLIT and NTRK-like family, member 6	axonogenesis
slit homolog 2 (Drosophila)	multicellular organismal development, cell differentiation
slit homolog 3 (Drosophila)	organ morphogenesis, multicellular organismal development, cell differentiation
solute carrier organic anion transporter family, member 4C1	multicellular organismal development, cell differentiation
sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1	multicellular organismal development
sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	regulation of cell differentiation
steroid sulfatase (microsomal), isozyme S	epidermis development
steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)	cell differentiation
steroid-5-alpha-reductase, alpha polypeptide 2 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 2)	cell differentiation
transforming growth factor, beta 1	odontogenesis, salivary gland morphogenesis, skeletal system development, lymph node development, regulation of cell differentiation, foam cell differentiation, regulation of transforming growth factor beta receptor signaling pathway
transforming growth factor, beta 2	odontogenesis, cell morphogenesis, salivary gland morphogenesis, angiogenesis, skeletal system development, neuron development, blood vessel development, transforming growth factor beta receptor signaling pathway
transforming growth factor, beta 3	odontogenesis, salivary gland morphogenesis
transmembrane protein with EGF-like and two follistatin-like domains 1	multicellular organismal development
tumor necrosis factor (ligand) superfamily, member 11	organ morphogenesis, multicellular organismal development, cell differentiation
tumor necrosis factor (ligand) superfamily, member 12	multicellular organismal development, cell differentiation
tumor necrosis factor (TNF superfamily, member 2)	organ morphogenesis, multicellular organismal development, positive regulation of hair follicle development, regulation of osteoclast differentiation
tumor necrosis factor receptor superfamily, member 11b	skeletal system development

Gene Name	Mapped Gene Ontology Term
vascular endothelial growth factor A	angiogenesis, multicellular organismal development
vascular endothelial growth factor C	organ morphogenesis, angiogenesis, multicellular organismal development, cell differentiation
v-Ha-ras Harvey rat sarcoma viral oncogene homolog	organ morphogenesis

Table 3

Over-represented developmental Gene Ontology terms that are unique to each class.
G&D= growth and development

Teratogen Risk Class	Total Unique Gene Ontology Terms	Highly Represented Processes/Systems (% of Total Terms)
A	0	None
B	45	Central Nervous system (22%) General G&D/embryogenesis (20%) Angiogenesis (11%)
C	91	General G&D/embryogenesis (18%) Musculoskeletal (15%) Immune (14%) Cutaneous (13%)
D	15	Cardiac/Vascular (27%) Immune (20%) General G&D/embryogenesis (13%) Musculoskeletal (13%)
X	10	General G&D/embryogenesis (60%) Gastrointestinal (40%)