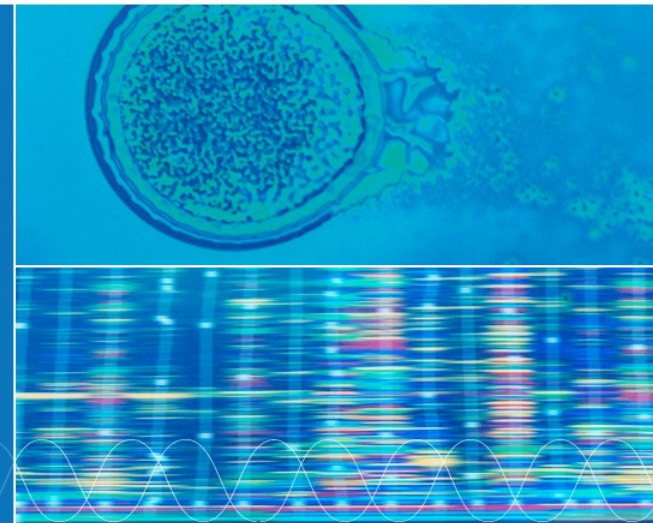




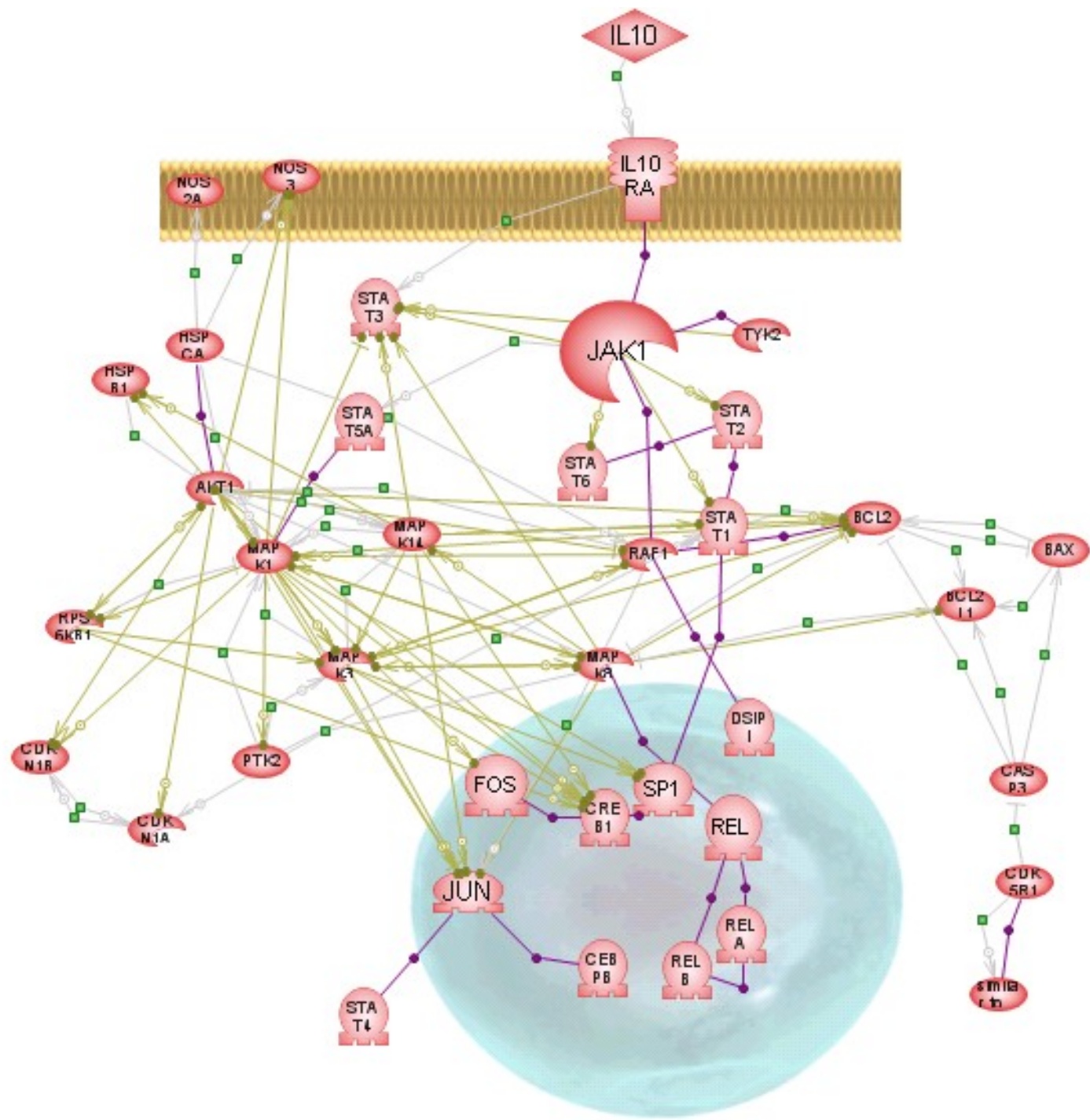
Are Pictures worth a Thousand Words?

Text Mining, Gisting & Visualisation

Martin Griffies M.Sc | October 2008



©2008 Ariadne Genomics. All Rights Reserved.



Agenda

- Information flow
- Gisting & visualisation
- Why visualisation helps
- Examples of visualisation

Text information flow & volumes

- 75,000 PubMed abstracts per month
- 45,000 patent *applications* per month
- 15,000 pharma-bio-med articles / m.
- Plus news feeds, internal sources, emails, blogs, etc.

Selection of relevant information

- Keyword search (on what?)
- Semantic search
- Alert services, review journals
- Serendipity

• *But it's still too much, because -*

SUMMARY OF THE INVENTION

[0014] The present inventors have found that variation in nicotine metabolism among individuals is due to variable expression of CYP2A isozymes; CYP2A6 has been shown to be the major nicotine metabolizing enzyme in human livers. Coumarin, a specific CYP2A6 substrate, was found to specifically and selectively inhibit nicotine metabolism to cotinine by 84%+-11% in test livers, and addition of orphenadrine (a CYP2B6 inhibitor) enhanced the inhibition. Methoxsalen and tranlylcypromine have also been found to be potent inhibitors of CYP2A6 and thus of nicotine to cotinine metabolism. The data indicate that variability in CYP2A6 expression results in inter-individual variation in nicotine metabolism, which in turn, can have behavioural consequences such as smoking more or less cigarettes. Therefore, inhibitors of CYP2A6 can be used to regulate nicotine metabolism, and in particular substantially decrease nicotine metabolism, thereby affecting tobacco use.

[0015] Broadly stated, the present invention relates to the diagnosis, prophylaxis and treatment of conditions requiring a reduction in the activity of a human cytochrome P450 enzyme CYP2A (referred to as "CYP2A" for brevity). The term "CYP2A" as used herein means all isoforms of CYP2A including but not limited to CYP2A(CYP1), CYP2A6, CYP2A7, CYP2A12, CYP2A13 and CYP2A16. Preferably the enzyme is CYP2A6.

[0016] The inventors have determined that the presence in an individual of a mutant allele of human cytochrome P450 enzyme CYP2A6 (referred to throughout this specification as "CYP2A6" for brevity) is predictive of an individual who: (i) has a decreased risk of becoming a smoker, (ii) will smoke less if he/she becomes dependent, and/or (iii) may be at relatively lower risk for cancer due to both decreased smoke exposure and decreased CYP2A6-mediated activation of tobacco smoke and other procarcinogenic substrates.

[0017] In one embodiment, this invention provides a diagnostic method for tobacco dependence risk and for cancers related to CYP2A6 substrates in an individual by analysing a DNA-containing bodily sample from the individual for the presence of a mutant allele of human cytochrome P450 enzyme CYP2A6. Preferably this method comprises genotype assaying the bodily sample, which may be genomic DNA isolated from peripheral leukocytes in the bodily sample. Alternatively the method comprises phenotype assaying the bodily sample, which may be a fluid, such as a blood sample or blood plasma. This invention also provides diagnostic kits for use in the analysis. The invention also provides a diagnostic method for tobacco dependence risk and for cancers related to human cytochrome P450 enzyme CYP2A6 substrates in an individual by administering a dose of a CYP2A6 substrate to the individual and determining in a bodily sample from the individual the level of said CYP2A6 substrate or a metabolite of said CYP2A6 substrate.

[0018] The invention specifically demonstrates that individuals who carry CYP2A6 deficient alleles are less likely to become smokers and will smoke less cigarettes if tobacco-dependent. In addition, because CYP2A6 is known to activate procarcinogens, such as those found in tobacco-smoke, the diagnostic aspect of the invention will be useful for identifying the contribution of this polymorphic locus to the genetic risk of an individual for cancer.

1 patent

24 pages

12,000 words

Reading & Comprehension

- Typical reading speed: 200 wpm
- Comprehension rate: 60%
- 1 scientific paper, 6000 words, 1/2 day.
- *Reading is not enough, so -*

Gisting / Summarisation

- Extraction of key terms & facts from corpus
- Improves first-pass comprehension
- Different from abstracts / reviews.
- Depends upon NLP technology (or humans)

Gate

- Applications
 - AMBIT
- Language Resources
 - text
 - GATE corpus_00049
- Processing Resources
 - Template Writer
 - Discourse Interpreter
 - Full Sentence Parser
 - POS Tagger
 - Term Orthographic Cor
 - Terminology Parser
 - CLEF Tags Annotator
 - PASTA test TM gazette
 - Ad Hoc Term Matcher
 - CLEF Lookup Faker
 - CLEF Tags Annotator
 - UMLS Annotator
 - CLEF Term Matcher

Messages GATE corpus_00049 text AMBIT

Text Annotations Annotation Sets Print

22.10.1998 **Seen** in Dr X's Clinic (MGTH).

Thank you very much indeed for asking us to **see** this delightful man who has had a **lymph node biopsy** which shows **melanoma** in his right **groin** which is clearly secondaries

from the **melanoma** on left **fourth toe**.

Although his **PET scan** is normal he does need a **groin dissection**. I **think** this is going to be quite difficult in view of his **infection** but we **will** perform a **CT scan** to look at the right pelvic side wall and I **will review** him together with X next week.

Clearly, he is a candidate for one of our current adjuvant **studies** and I **will** write to you next week with a more detailed **treatment plan**.

Thank you very much indeed for asking us to become involved with his care.

Very best wishes

Letter to Dr X from Dr Y, PhD FRCP

Dr X/RAJ/26.10.98/13:55

- Coreference
- DBInstance
- Lookup
- Muc_coref
- Name
- Semantics
- Sentence
- SentenceSem
- SpaceToken
- Split
- Syntax
- Template
- Template_Relation
- Token
- UMLSTerm
- Xi_instance
- Xi_node
- clefAnatLocus
- clefInterv
- clefInvest
- clefNonAnatLocus
- clefProblem

Original markups annotati

Annotations Editor Features Editor Initialisation Parameters

The expression and prognostic value of the guanine nucleotide exchange factors (GEFs) Trio, Vav1 and TIAM-1 in human breast cancer

in *Surgical Oncology*
Volume 5

Viewing options:

Jane Lane , Tracey A Martin , Robert E Mansel  and Wen G Jiang 

International Seminars in Surgical Oncology 2008, **5**:23 doi:10.1186/1477-7800-5-

Published: 16 October 2008

[Abstract \(provisional\)](#)

Background

Development of metastasis in breast cancer is a multi-step process comprising changes in cytoskeletal structure and gene expression of tumour cells leading to changes in cell adhesion and motility. The Rho-GTPases, governed by guanine nucleotide regulated binary switches, govern a variety of cellular processes in cell adhesion as well as actin cytoskeletal reorganisation and gene expression/transcription. The Rho-GTPases is the guanine nucleotide exchange factors (GEFs), and this study investigated the expression of these factors (Trio, Vav1 and TIAM-1). The purpose of this study was to investigate the expression of these factors on clinical outcome.

Methods

Specimens of fresh, frozen breast tumour tissue (n=113) and normal background breast tissue were analysed by RT-PCR analysis. The expression and levels of expression of Trio, Vav1 and TIAM-1 were determined respectively. Sections were also immunostained with Trio and Tiam-1 antibodies.

Results

Tumour tissue exhibited high levels of all three Rho activators Trio, Vav1 and TIAM-1 compared with normal background breast tissue, reaching a level of significance for the GEF Trio ($p=0.013$). Trio levels also increased significantly in patients with a poor prognostic index ($p=0.04$). Levels of TIAM-1 were significantly higher in tumour tissue from patients who died from breast cancer compared with those who survived ($p=0.04$). No significant correlation was found between tumour grade and histology types.

Conclusions

High expression levels of Trio, Vav1 and TIAM-1 were seen in breast tumours, especially in those with poor prognosis. This suggests that aberrant regulation of Rho family activities by GEFs may have an important prognostic value in breast cancer.

QuickJist - Web Page Summary

Development of metastasis in breast cancer is a multi-step process comprising changes in cytoskeletal structure and gene expression of tumour cells leading to changes in cell adhesion and motility.

High expression levels of Trio, Vav1 and TIAM-1 were seen in breast tumours, especially in those with poor prognosis.

Copy summary to clipboard

Options...

Close

ting digtext:C:\Documents and Settings\Martin\Desktop\Demonstration\Current_Ariadne_test080820.txt

ss: Processing Citations: read 2, w/ab 2, w/mnos 0, processed 2 Citation duplicates: 0 Relevant citations: 1=50.0% of processed
 ces processed: 740 (329=44.5% of them relevant) Time: 0.32 sec (6.27 rps / 6.27 cps) Elapsed time (locscan): 0.28 secs Totals: 329 input
 ces, 1239 marked-up entities. Elapsed time: 0.28 secs

ected sentences

- Sentence 1 1: Phytother Res. 2008 Aug 7. [Epub ahead of print] Related Articles, Links Click here to read Resveratrol induces apoptosis and inhibits adipogenesis in 3T3-L1 adipocytes.
- Sentence 6 In this study, we investigated the effects of resveratrol on adipogenesis and apoptosis using 3T3-L1 cells.
- Sentence 7 In mature adipocytes, 100 and 200 microM resveratrol decreased cell viability dose-dependently by 23 +/- 2.7%, and 75.3 +/- 2.8% (p < 0.0001), respectively, after 48 h treatment, and 100 microM resveratrol increased apoptosis by 76 +/- 8.7% (p < 0.0001).
- Sentence 8 Resveratrol at 25 and 50 microM decreased lipid accumulation in maturing preadipocytes significantly by 43 +/- 1.27% and 94.3 +/- 0.3% (p < 0.0001) and decreased cell viability by 25 +/- 1.3% and 70.4 +/- 1.6% (p < 0.0001), respectively.
- Sentence 9 In order to understand the anti-adipogenic effects of resveratrol, maturing 3T3-L1 preadipocytes were treated with 25 microM resveratrol and the change in the expression of several adipogenic transcription factors and enzymes was investigated using real-time RT-PCR.
- Sentence 10 Resveratrol down-regulated the expression of PPARgamma, C/EBPalpha, SREBP-1c, FAS, HSL, LPL and up-regulated the expression of genes regulating mitochondrial activity (SIRT3, UCP1 and Mfn2).
- Sentence 11 These results indicate that resveratrol may alter fat mass by directly affecting cell viability and adipogenesis in maturing preadipocytes and inducing apoptosis in adipocytes and thus may have applications for the treatment of obesity.
- Sentence 14 2: Circ Res. 2008 Aug 7. [Epub ahead of print] Related Articles, Links Click here to read Phosphatidylinositol 3-Kinase {gamma} Is a Critical Mediator of Myocardial Ischemic and Adenosine-Mediated Preconditioning.
- Sentence 17 Ischemic preconditioning (IPC) is a potent cellular protective mechanism whereby brief periods of sublethal ischemia protect the myocardium from prolonged ischemia-induced injury.
- Sentence 18 We demonstrate the selective role of phosphatidylinositol 3-kinase (PI3K) isoforms in IPC.
- Sentence 20 Examination of the cell-signaling pathways revealed restored phosphorylation levels of Akt and glycogen synthase kinase (GSK)3beta in wild-type hearts, which were abolished in PI3Kgamma(-/-) hearts subjected to IPC.
- Sentence 21 Inhibition of GSK3beta by LiCl reversed the loss in protection in PI3Kgamma(-/-) hearts.
- Sentence 22 In contrast, mice expressing a cardiac-specific kinase-deleted PI3Kalpha (PI3KalphaDN) were resistant to injury induced by 30 minutes of ischemia followed by 40 minutes of reperfusion.
- Sentence 23 Furthermore, the resistance of PI3KalphaDN hearts to ischemia/reperfusion correlated with the persistent expression of p110gamma and was blocked by the PI3K inhibitor wortmannin, suggesting the possible enhanced cell signaling through the PI3Kgamma pathway.

Current Page Digest

- 11 PRKDC
- 10 isoflavone
- 10 FOXO3A
- 10 GSK3B
- 10 inflammation
- 10 lung cancer
- 9 HSPC150
- 9 radiation
- 8 IL6
- 8 asthma
- 8 alpha-methylstyrene
- 8 amentoflavone
- 8 cell survival
- 8 EGFR
- 8 small nuclear ribonucleoprotein

124 Relations 1031 Entities

Known Relations

- 1 resveratrol → Sirt3
- 1 PRPF4B → mitosis
- 1 PRPF4B → mRNA splicing
- 1 PRPF4B → CDC2L1
- 1 genistein → AKT1
- 1 isoflavone → AKT1
- 1 isoflavone → FOXO3A
- 1 isoflavone → SRC
- 1 isoflavone → SRC
- 1 BIRC5 → apoptosis

Gisting/ Summarisation

Relations	
108	CYP2A6
58	nicotine
32	CYP2A
26	nicotine metabolism
23	CYP2B6
21	methoxsalen
20	coumarin
18	cotinine
16	cancer
14	tranylcypromine
10	monooxygenase
8	orphenadrine
6	microsome
6	Hypericum
5	Coumarin 7
5	nucleic acid
4	transdermal
4	Cichorium intybus
4	sucrose
3	Fatty acid complexes
3	carbohydrates
3	sodium salts
18 Relations	259 Entities

SUMMARY OF THE INVENTION

[0014] The present inventors have found that variation in nicotine metabolism among individuals is due to variable expression of CYP2A isozymes; CYP2A6 has been shown to be the major nicotine metabolizing enzyme in human livers. Coumarin, a specific CYP2A6 substrate, was found to specifically and selectively inhibit nicotine metabolism to cotinine by 84%+-11% in test livers, and addition of orphenadrine (a CYP2B6 inhibitor) enhanced the inhibition. Methoxsalen and tranylcypromine have also been found to be potent inhibitors of CYP2A6 and thus of nicotine to cotinine metabolism. The data indicate that variability in CYP2A6 expression results in inter-individual variation in nicotine metabolism, which in turn, can have behavioural consequences such as smoking more or less cigarettes. Therefore, inhibitors of CYP2A6 can be used to regulate nicotine metabolism, and in particular substantially decrease nicotine metabolism, thereby affecting tobacco use.

[0015] Broadly stated, the present invention relates to the diagnosis, prophylaxis and treatment of conditions requiring a reduction in the activity of a human cytochrome P450 enzyme CYP2A (referred to as "CYP2A" for brevity). The term "CYP2A" as used herein means all isoforms of CYP2A including but not limited to CYP2A(CYP1), CYP2A6, CYP2A7, CYP2A12, CYP2A13 and CYP2A16. Preferably the enzyme is CYP2A6.

[0016] The inventors have determined that the presence in an individual of a mutant allele of human cytochrome P450 enzyme CYP2A6 (referred to throughout this specification as "CYP2A6" for brevity) is predictive of an individual who: (i) has a decreased risk of becoming a smoker, (ii) will smoke less if he/she becomes dependent, and/or (iii) may be at relatively lower risk for cancer due to both decreased smoke exposure and decreased CYP2A6-mediated activation of tobacco smoke and other procarcinogenic substrates.

[0017] In one embodiment, this invention provides a diagnostic method for tobacco dependence risk and for cancers related to CYP2A6 substrates in an individual by analysing a DNA-containing bodily sample from the individual for the presence of a mutant allele of human cytochrome P450 enzyme CYP2A6. Preferably this method comprises genotype assaying the bodily sample, which may be genomic DNA isolated from peripheral leukocytes in the bodily sample. Alternatively the method comprises phenotype assaying the bodily sample, which may be a fluid, such as a blood sample or blood plasma. This invention also provides diagnostic kits for use in the analysis. The invention also provides a diagnostic method for tobacco dependence risk and for cancers related to human cytochrome P450 enzyme CYP2A6 substrates in an individual by administering a dose of a CYP2A6 substrate to the individual and determining in a bodily sample from the individual the level of said CYP2A6 substrate or a metabolite of said CYP2A6 substrate.

[0018] The invention specifically demonstrates that individuals who are carry CYP2A6 deficient alleles are less likely to become smokers and will smoke less cigarettes if tobacco-dependent. In addition, because CYP2A6 is known to activate procarcinogens, such as those found in tobacco-smoke, the diagnostic aspect of the invention will be useful for identifying the contribution of this polymorphic locus to the genetic risk of an individual for cancer.

digtext:C:\Documents and Settings\Martin\Desktop\Demonstration\Biberach_pubmed-result.txt

processing ... Citations: read 2, w/ab 2, w/mnos 0, processed 2 Citation duplicates: 0 Relevant citations: 1=50.0% of processed Sentences
 4536 (1232=27.2% of them relevant) Time: 20.24 sec (0.10 rps / 0.10 cps) Elapsed time (locscan): 21.34 secs Totals: 1232 input sentences, 4193
 entities. Elapsed time: 21.05 secs

Sentences

from C:\Documents and Settings\Martin\Desktop\Demonstration\Biberach_pubmed-result.txt

Sentence 2 In-vitro profile and ex-vivo anticoagulant activity of the direct thrombin inhibitor dabigatran and its orally active prodrug, dabigatran etexilate.

Sentence 6 Dabigatran is a reversible and selective, direct thrombin inhibitor (DTI) undergoing advanced clinical development as its orally active prodrug, dabigatran etexilate.

Sentence 7 This study set out to determine the molecular potency and anticoagulant efficacy of dabigatran and its prodrug dabigatran etexilate.

Sentence 8 This was achieved through enzyme inhibition and selectivity analyses, surface plasmon resonance studies, platelet aggregation, thrombin generation and clotting assays in vitro and ex vivo.

Sentence 9 These studies demonstrated that dabigatran selectively and reversibly inhibited human thrombin (K_i: 4.5 nM) as well as thrombin-induced platelet aggregation (IC₅₀: 10 nM), while showing no inhibitory effect on other platelet-stimulating agents.

Sentence 10 Thrombin generation in platelet-poor plasma (PPP), measured as the endogenous thrombin potential (ETP) was inhibited concentration-dependently (IC₅₀: 0.56 microM).

Sentence 11 Dabigatran demonstrated concentration-dependent anticoagulant effects in various species in vitro, doubling the activated partial thromboplastin time (aPTT), prothrombin time (PT) and ecarin clotting time (ECT) in human PPP at concentrations of 0.23, 0.83 and 0.18 microM, respectively.

Sentence 13 Dose- and time-dependent anticoagulant effects were observed with dabigatran etexilate administered orally to conscious rats (10, 20 and 50 mg/kg) or rhesus monkeys (1, 2.5 or 5 mg/kg), with maximum effects observed between 30 and 120 min after administration, respectively.

Sentence 14 These data suggest that dabigatran is a potent, selective thrombin inhibitor and an orally active anticoagulant as the prodrug, dabigatran etexilate.

Sentence 22 Aortic pressure (AP), left ventricular pressure (LVP), a lead II ECG and body temperature could be continuously monitored.

Sentence 23 The contractility index LVdP/dt_{max} was derived from the LVP signal.

Sentence 25 For each species an LVdP/dt-heart rate relationship was evaluated using spontaneous heart rates (HR) throughout the observation period.

Sentence 26 A validation compound with positive inotropic effects (pimobendan) was then used to investigate the LVdP/dt-heart rate relationship.

Sentence 29 Discussion: Contractility of the myocardium is regulated by autonomic input activating primarily myocardial beta1-adrenoceptors, but it is also affected by the "force-frequency" relationship.

Regulation: Discussion: Contractility of the myocardium is regulated by autonomic input activating primarily myocardial beta1-adrenoceptors, but it is also affected by the "force-frequency" relationship.

Current Page Digest

1	acetylcholine	→	cell proliferation
1	acetylcholine	→	relaxation
1	ADM	→	vasodilation
1	adrenoceptor	→	contraction
1	AGT	→	blood pressure

1	AGT	→	hypertension
1	ALB	→	kidney excretion
1	alzheimer disease	→	proteasome
1	anticoagulant	→	thrombosis
1	APP	→	Glucose metabolism
1	ATP2A2	→	contraction
1	beta adrenoceptor	→	heart rate
1	beta-galactosidase	↔	FAP
1	Bibn-99	→	CHRM2
1	BIBX 79	→	oxidosqualene cy...
1	CALCA	→	Ischemia
1	CALCA	→	ganglion stimulation
1	CALCA	→	migraine
1	CALCA	→	vasodilation
1	CALCA	→	blood flow

172 Relations / 1101 Entities

Known Relations

Parse this, Jimmie.....

- and **in its absence, deficient 60 S ribosomes are assembled** which are inactive in protein synthesis resulting in cell lethality.
- Mutations that completely abolish recognition of 26 S rRNA, however, block the formation of 60S particles, demonstrating that **binding of L25 to this rRNA is an essential step in the assembly of the large ribosomal subunit.**
- -**Depletion of Saccharomyces cerevisiae ribosomal protein L16 causes decrease in 60S ribosomal subunits** and formation of half-mer polyribosomes.
- Without L3, apparent synthesis of several 60 S subunit proteins diminished, and 60S subunit did not assemble. A similar phenomenon occurred, when a second strain, **synthesis of ribosomal protein L29 was prevented.**

Ribosome assembly and maintenance

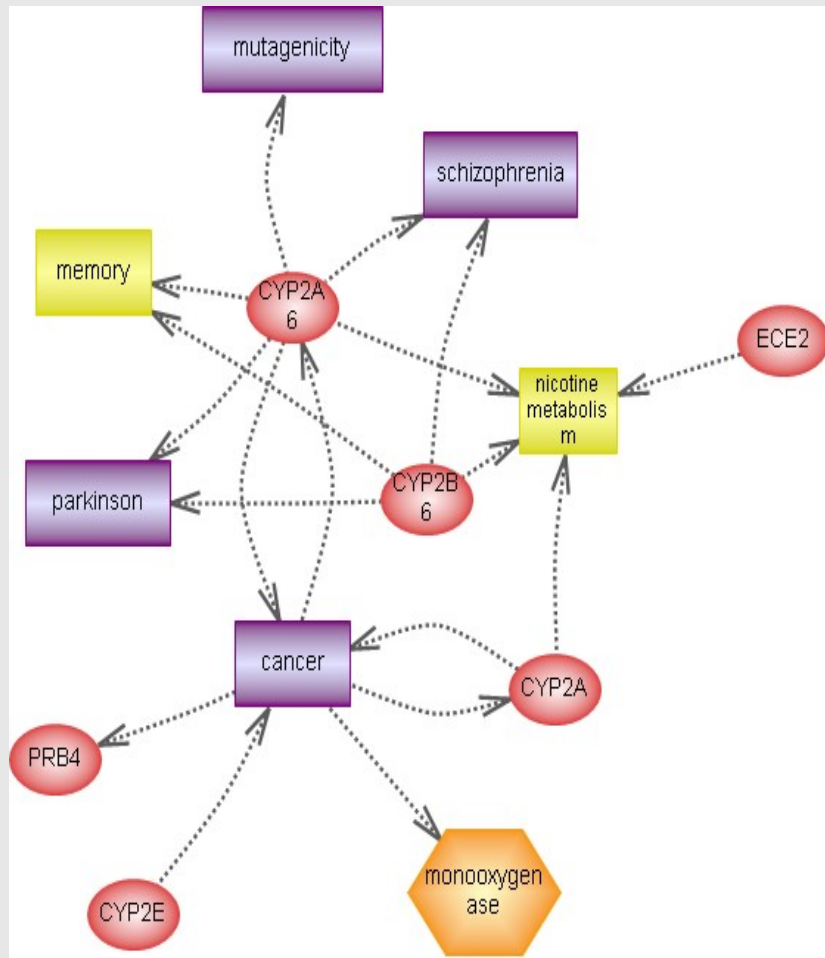
Gisting requirements

- Good entity recognition
 - Dictionary-based
 - Inferential-based
- Domain specificity
- Linguistic analysis for fact extraction
- Extensible ontology & dis-ambiguation
- (Use of metadata)

Gisting / Summarisation Overview

- Fast and accurate (with good algorithms)
- Works with corpora or single documents
- Improves rapid understanding.
- Entities and facts extracted
- Provides information from full-text
 - (Compare with abstracts)
 - (2-4 times as many FACTS)

Visualisation



SUMMARY OF THE INVENTION

[0014] The present inventors have found that variation in nicotine metabolism among individuals is due to variable expression of CYP2A isozymes; CYP2A6 has been shown to be the major nicotine metabolizing enzyme in human livers. Coumarin, a specific CYP2A6 substrate, was found to specifically and selectively inhibit nicotine metabolism to cotinine by 84%+-11% in test livers, and addition of orphenadrine (a CYP2B6 inhibitor) enhanced the inhibition. Methoxsalen and tranylcypromine have also been found to be potent inhibitors of CYP2A6 and thus of nicotine to cotinine metabolism. The data indicate that variability in CYP2A6 expression results in inter-individual variation in nicotine metabolism, which in turn, can have behavioural consequences such as smoking more or less cigarettes. Therefore, inhibitors of CYP2A6 can be used to regulate nicotine metabolism, and in particular substantially decrease nicotine metabolism, thereby affecting tobacco use.

[0015] Broadly stated, the present invention relates to the diagnosis, prophylaxis and treatment of conditions requiring a reduction in the activity of a human cytochrome P450 enzyme CYP2A (referred to as "CYP2A" for brevity). The term "CYP2A" as used herein means all isoforms of CYP2A including but not limited to CYP2A(CYP1), CYP2A6, CYP2A7, CYP2A12, CYP2A13 and CYP2A16. Preferably the enzyme is CYP2A6.

[0016] The inventors have determined that the presence in an individual of a mutant allele of human cytochrome P450 enzyme CYP2A6 (referred to throughout this specification as "CYP2A6" for brevity) is predictive of an individual who: (i) has a decreased risk of becoming a smoker, (ii) will smoke less if he/she becomes dependent, and/or (iii) may be at relatively lower risk for cancer due to both decreased smoke exposure and decreased CYP2A6-mediated activation of tobacco smoke and other procarcinogenic substrates.

[0017] In one embodiment, this invention provides a diagnostic method for tobacco dependence risk and for cancers related to CYP2A6 substrates in an individual by analysing a DNA-containing bodily sample from the individual for the presence of a mutant allele of human cytochrome P450 enzyme CYP2A6. Preferably this method comprises genotype assaying the bodily sample, which may be genomic DNA isolated from peripheral leukocytes in the bodily sample. Alternatively the method comprises phenotype assaying the bodily sample, which may be a fluid, such as a blood sample or blood plasma. This invention also provides diagnostic kits for use in the analysis. The invention also provides a diagnostic method for tobacco dependence risk and for cancers related to human cytochrome P450 enzyme CYP2A6 substrates in an individual by administering a dose of a CYP2A6 substrate to the individual and determining in a bodily sample from the individual the level of said CYP2A6 substrate or a metabolite of said CYP2A6 substrate.


[0018] The invention specifically demonstrates that individuals who carry CYP2A6 deficient alleles are less likely to become smokers and will smoke less cigarettes if tobacco-dependent. In addition, because CYP2A6 is known to activate procarcinogens, such as those found in tobacco-smoke, the diagnostic aspect of the invention will be useful for identifying the contribution of this polymorphic locus to the genetic risk of an individual for cancer.

Why Use Visualisation?

- 50% of brain used for processing vision
- Difficult concepts simplified
- Rapid comprehension
- Multidimensional
- Serendipity

Tables / spreadsheets, versus

CONFIG



Barrow All
Barrow UNC
Industry Developments
Competitors
OC Customers
OC Orders
Bug Tracker
Franks Mail

employee_name

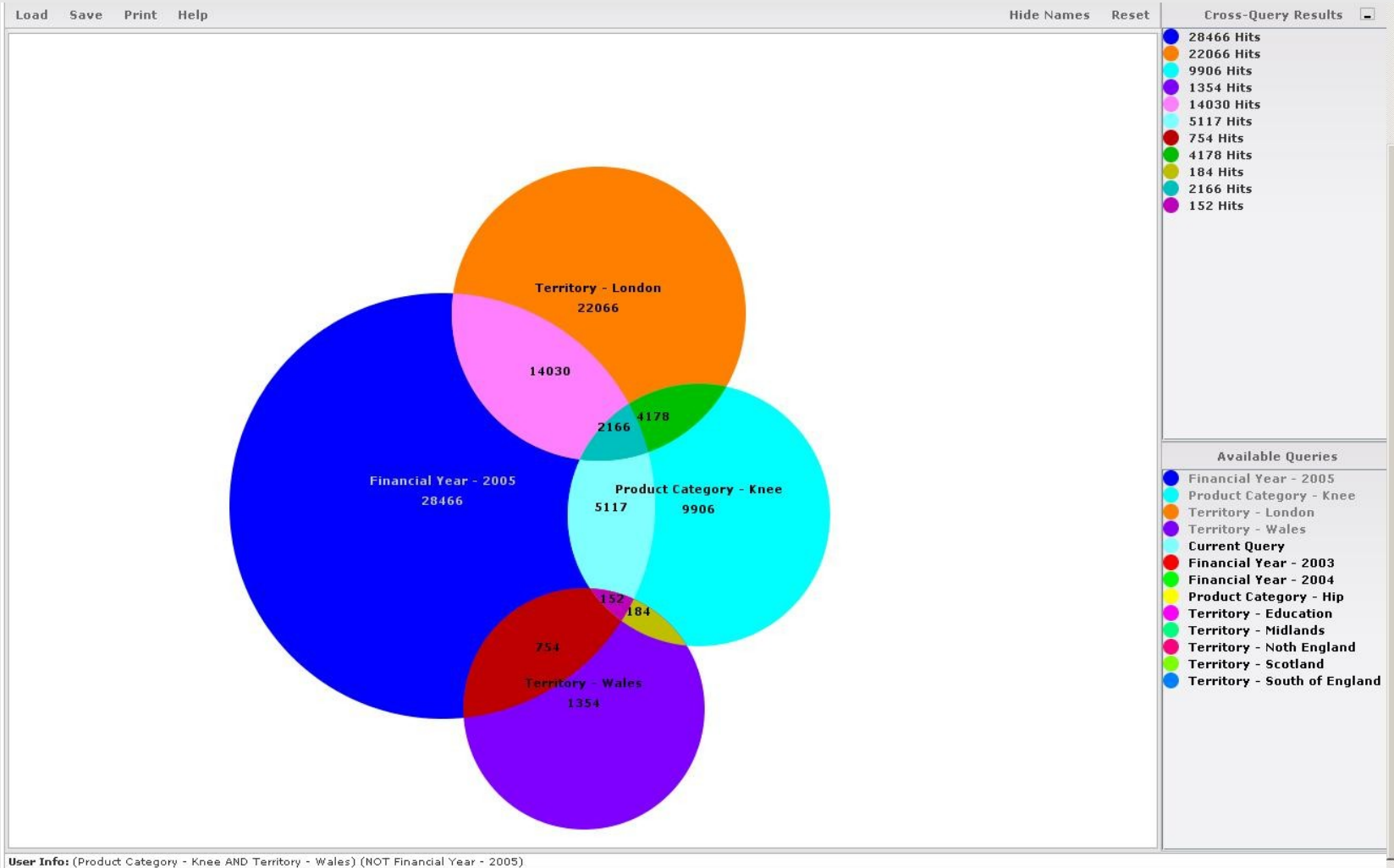
All Fields Search Clear ?

Combine

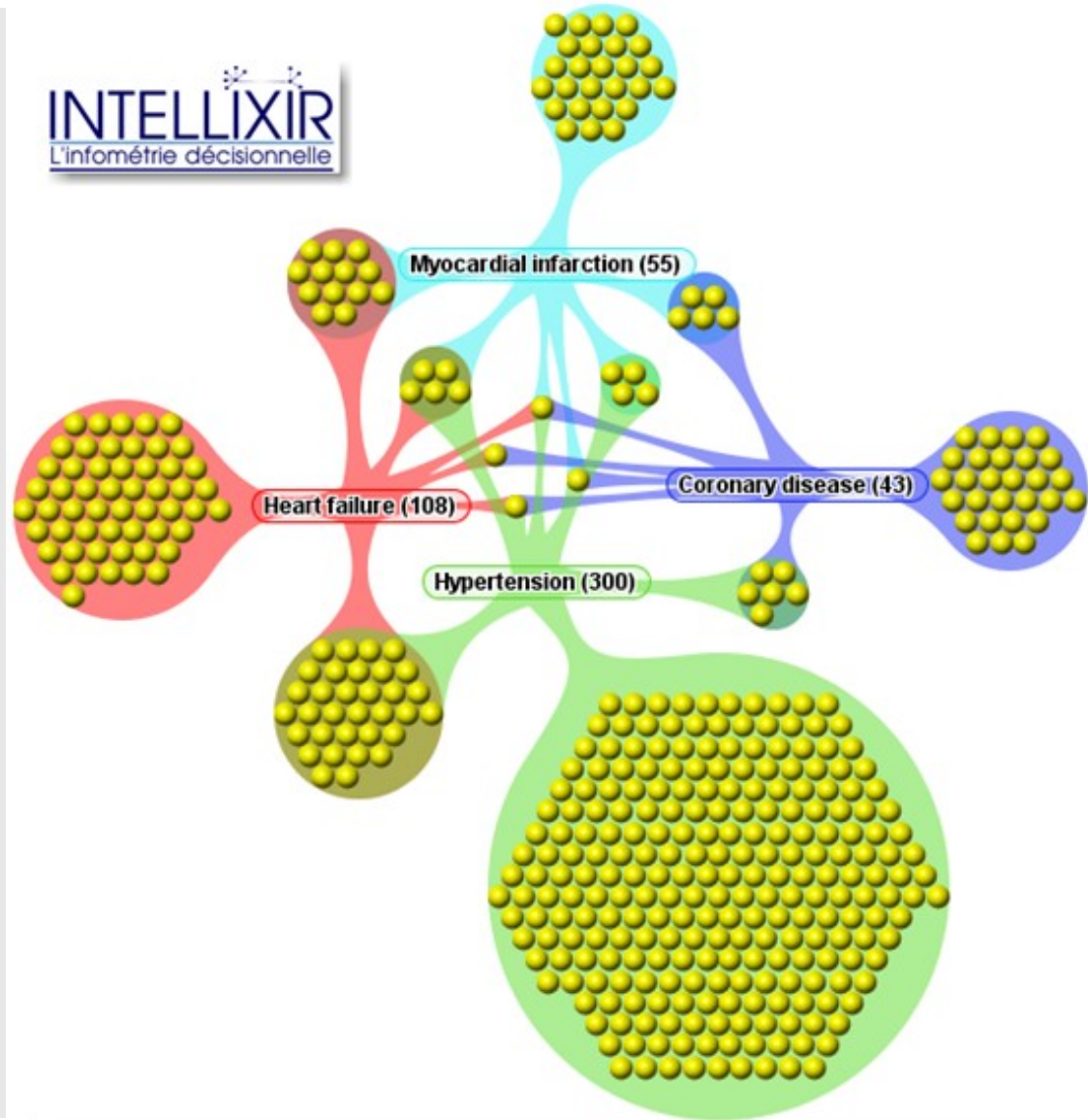
Queries

<input type="checkbox"/> bupa <input type="checkbox"/> Financial Year - 2003 <input type="checkbox"/> Financial Year - 2004 <input checked="" type="checkbox"/> Financial Year - 2005 <input checked="" type="checkbox"/> Product Category - Hip <input checked="" type="checkbox"/> Product Category - Knee <input type="checkbox"/> Territory - Education <input checked="" type="checkbox"/> Territory - London <input checked="" type="checkbox"/> Territory - Midlands <input checked="" type="checkbox"/> Territory - Noth England <input checked="" type="checkbox"/> Territory - Scotland <input checked="" type="checkbox"/> Territory - South of England <input checked="" type="checkbox"/> Territory - Wales	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="font-size: 8px; text-decoration: underline;">Financial Year - 2005</th> <th style="font-size: 8px; text-decoration: underline;">Product Category - Hip</th> <th style="font-size: 8px; text-decoration: underline;">Product Category - Knee</th> <th style="font-size: 8px; text-decoration: underline;">Territory - London</th> <th style="font-size: 8px; text-decoration: underline;">Territory - Midlands</th> <th style="font-size: 8px; text-decoration: underline;">Territory - Noth England</th> <th style="font-size: 8px; text-decoration: underline;">Territory - Scotland</th> <th style="font-size: 8px; text-decoration: underline;">Territory - South of England</th> <th style="font-size: 8px; text-decoration: underline;">Territory - Wales</th> </tr> </thead> <tbody> <tr> <td style="font-size: 8px; text-decoration: underline;">Financial Year - 2005</td> <td>50685</td> <td>37908</td> <td>7435</td> <td>16196</td> <td>6576</td> <td>9612</td> <td>536</td> <td>14220</td> <td>906</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Product Category - Hip</td> <td>37908</td> <td>97686</td> <td>0</td> <td>30662</td> <td>13710</td> <td>18797</td> <td>1008</td> <td>26623</td> <td>1820</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Product Category - Knee</td> <td>7435</td> <td>0</td> <td>21703</td> <td>6344</td> <td>3061</td> <td>4036</td> <td>252</td> <td>6386</td> <td>336</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Territory - London</td> <td>16196</td> <td>30662</td> <td>6344</td> <td>42440</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Territory - Midlands</td> <td>6576</td> <td>13710</td> <td>3061</td> <td>0</td> <td>19262</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Territory - Noth England</td> <td>9612</td> <td>18797</td> <td>4036</td> <td>0</td> <td>0</td> <td>26224</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Territory - Scotland</td> <td>536</td> <td>1008</td> <td>252</td> <td>0</td> <td>0</td> <td>0</td> <td>1490</td> <td>0</td> <td>0</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Territory - South of England</td> <td>14220</td> <td>26623</td> <td>6386</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>37792</td> <td>0</td> </tr> <tr> <td style="font-size: 8px; text-decoration: underline;">Territory - Wales</td> <td>906</td> <td>1820</td> <td>336</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>2444</td> </tr> </tbody> </table>	Financial Year - 2005	Product Category - Hip	Product Category - Knee	Territory - London	Territory - Midlands	Territory - Noth England	Territory - Scotland	Territory - South of England	Territory - Wales	Financial Year - 2005	50685	37908	7435	16196	6576	9612	536	14220	906	Product Category - Hip	37908	97686	0	30662	13710	18797	1008	26623	1820	Product Category - Knee	7435	0	21703	6344	3061	4036	252	6386	336	Territory - London	16196	30662	6344	42440	0	0	0	0	0	Territory - Midlands	6576	13710	3061	0	19262	0	0	0	0	Territory - Noth England	9612	18797	4036	0	0	26224	0	0	0	Territory - Scotland	536	1008	252	0	0	0	1490	0	0	Territory - South of England	14220	26623	6386	0	0	0	0	37792	0	Territory - Wales	906	1820	336	0	0	0	0	0	2444
Financial Year - 2005	Product Category - Hip	Product Category - Knee	Territory - London	Territory - Midlands	Territory - Noth England	Territory - Scotland	Territory - South of England	Territory - Wales																																																																																												
Financial Year - 2005	50685	37908	7435	16196	6576	9612	536	14220	906																																																																																											
Product Category - Hip	37908	97686	0	30662	13710	18797	1008	26623	1820																																																																																											
Product Category - Knee	7435	0	21703	6344	3061	4036	252	6386	336																																																																																											
Territory - London	16196	30662	6344	42440	0	0	0	0	0																																																																																											
Territory - Midlands	6576	13710	3061	0	19262	0	0	0	0																																																																																											
Territory - Noth England	9612	18797	4036	0	0	26224	0	0	0																																																																																											
Territory - Scotland	536	1008	252	0	0	0	1490	0	0																																																																																											
Territory - South of England	14220	26623	6386	0	0	0	0	37792	0																																																																																											
Territory - Wales	906	1820	336	0	0	0	0	0	2444																																																																																											

Visualisation....



Heat maps

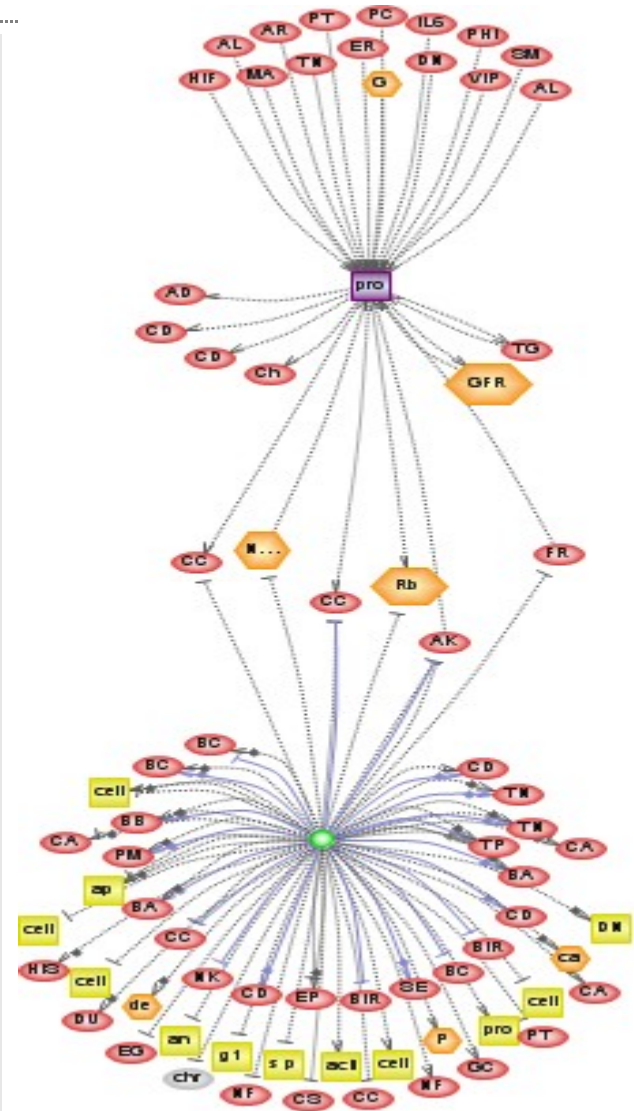


Visualisation & Network (Pathway) Analysis

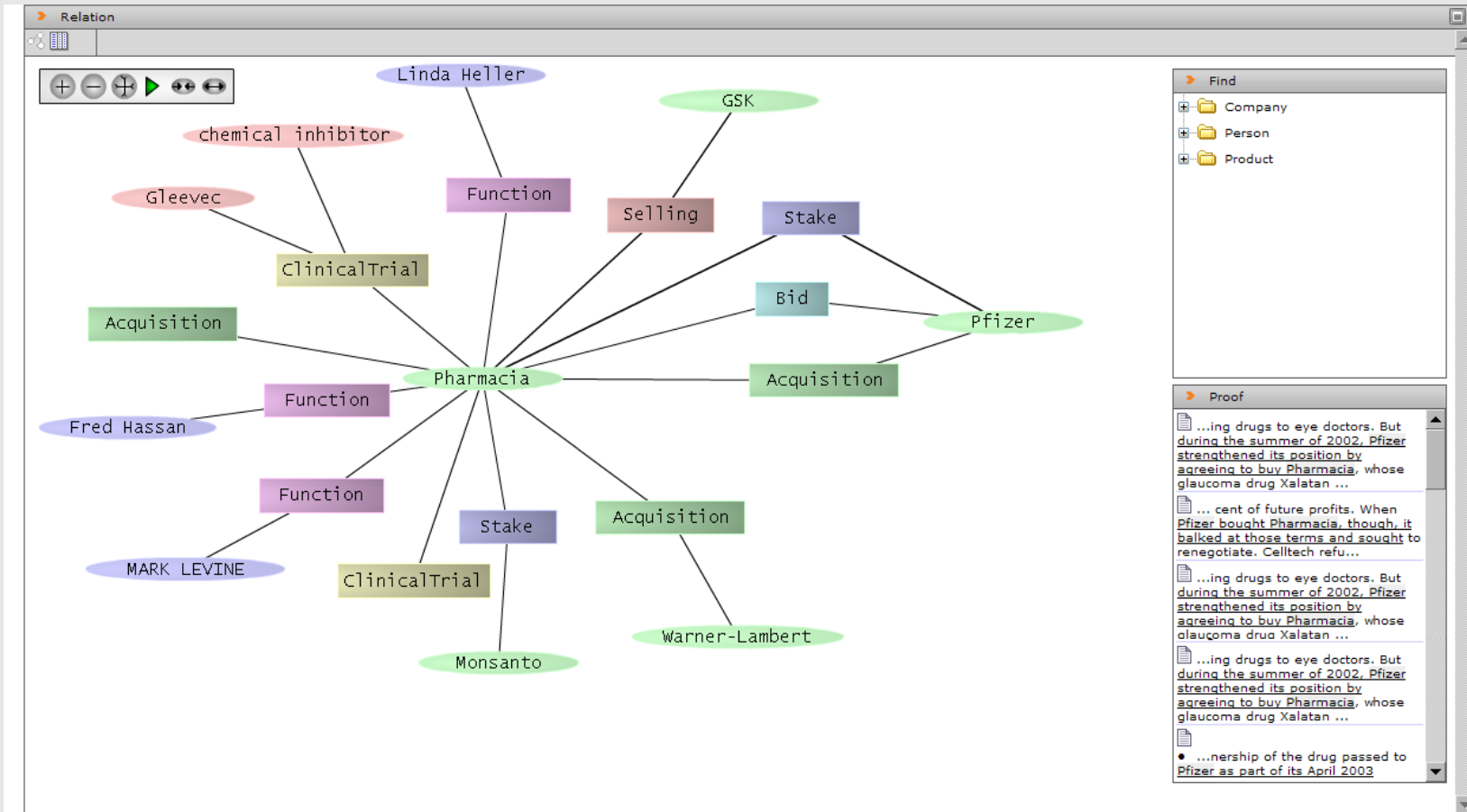
- Isolated facts from disparate sources
- Ontology-dependent for effectiveness
- Synthetic capability - joining the dots...
- Will need verification
 - Don't believe everything
- Provides depth of information

Visualisation in science

- 2 key words
- 84 abstracts
- 782 sentences
- 311 relationships / facts
- <5 minutes
- Corpus: >all PubMed



Visualisation in business



In summary....

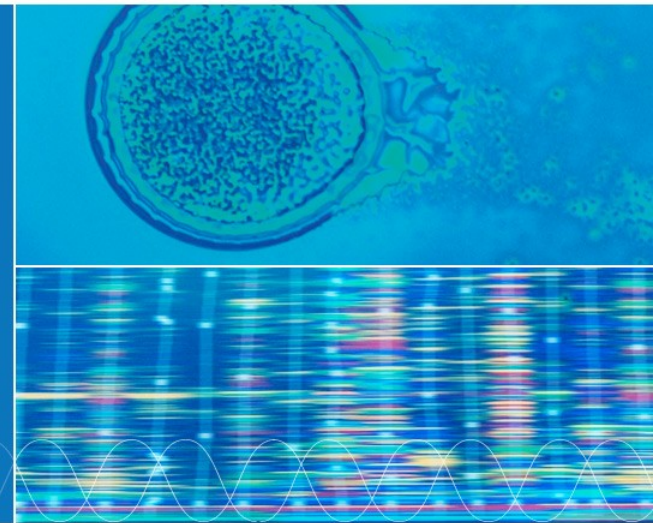
Gisting & visualisation are

- Effective end-user tools for understanding
- Professional tools for information reporting
- Tools for knowledge discovery
- Tools for information presentation

Thank you. Any questions?

Text Mining, Gisting & Visualisation

Martin Griffies M.Sc | mjg@ariadnegenomics.com



©2008 Ariadne Genomics. All Rights Reserved.

Acknowledgements: Nikolai Daraselia, Ariadne; Ilya Mazo, Ariadne; Vitaly Demin, QuickJist; Temis SA; Prof. Tsujii, NaCTeM; Henk Harkema, Sheffield Uni; Intellixir

Agenda

- Information flow
- Gisting & visualisation
- Why visualisation helps
- Examples of visualisation
- When visualisation is NOT helpful

NO BACKGROUND PATTERN

Slide Subtitle

- Lorem Ipsum Lorem Ipsum
 - Duis autem vel eum iriure dolor in hendrerit in vulputate
 - Lorem ipsum dolor sit amet, consectetur adipiscing
 - Vel eum iriure dolor
 - Ipsum dolor sit amet
 - In hendrerit in vulputate
- Vel Eum Iriure Dolor
- Ipsum dolor sit amet dolor sit amet
- Duis autem vel eum sit amet

