ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

OECD OCDE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

OECD WORKSHOP

An International Perspective on Pharmacogenetics: the intersection between innovation, regulation and health delivery

Rome, Italy, 17-19 OCTOBER, 2005

The druggable genome:

How do we deliver on the promise of new drug targets?

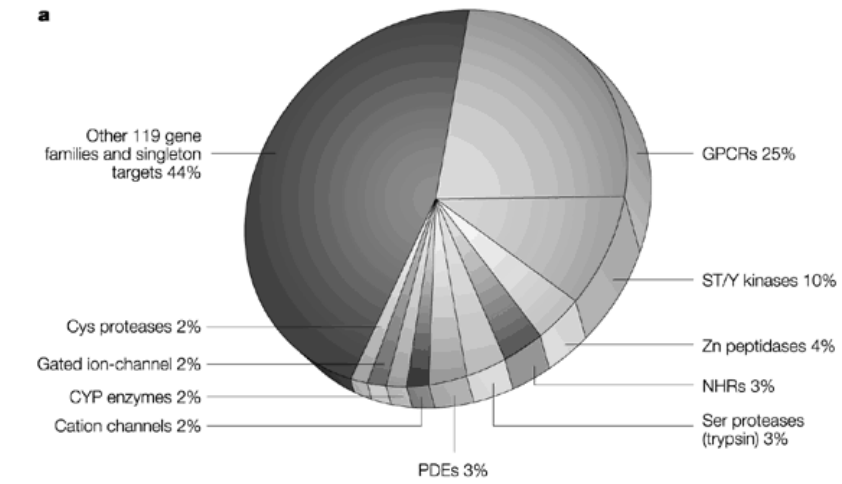
Andrew L. Hopkins

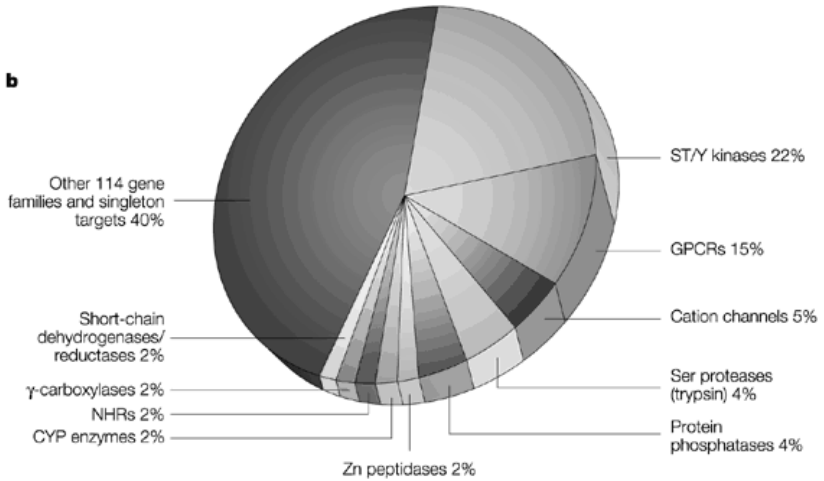
Sandwich Laboratories,

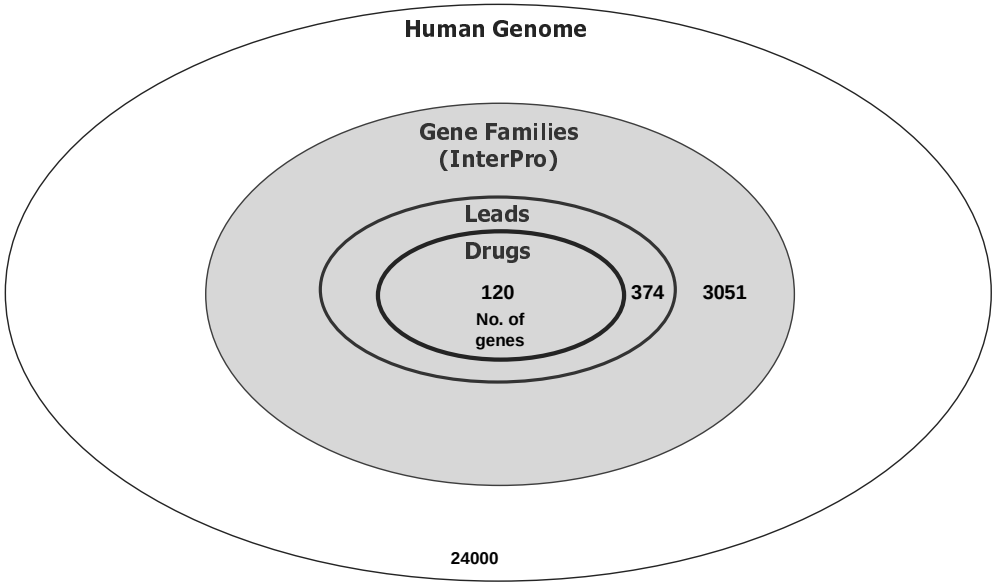
Pfizer Global Research & Development

andrew.hopkins@pfizer.com

Druggable Genome (2002)







- 399 targets (Ro5, >10μM)
- 129 Druggable gene families
- 13% human genome

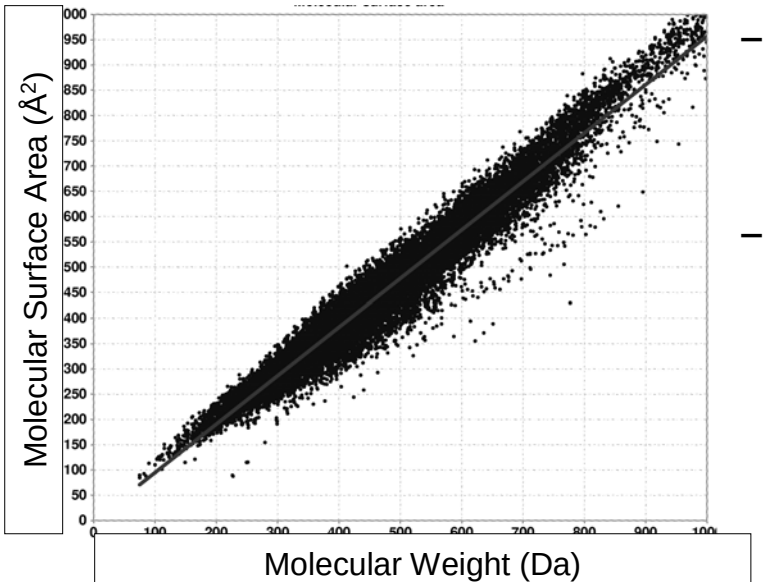
Table 1 | **Comparison of the druggable genomes of selected eukaryotes**

	Homo sapiens	*Drosophila melanogaster*	*Caenorhabditis elegans*
Total number of predicted genes[8,9,16]	~30,000	13,601	18,424
Number of proteins in proteome*	21,688	13,849	17,946
Number of estimated druggable targets	3,051	1,714	2,267
Percentage that are predicted druggable targets	~10–14%	12%	12%

Hopkins & Groom, Nature Rev Drug Disc. (2002), 1 (9), 727-730

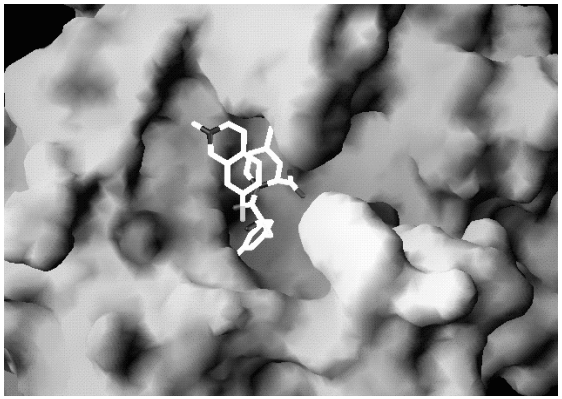
Basic Physico-chemical limits

- Binding is determined by buried surface area contacts and polar interactions
  - $\Delta G = -RT \ln (K_i)$
    - Every 10-fold increase in potency = -1.36 kcal/mol
    - K_i of 10nM = -11 kcal/mol
  - The free energy gained from burying
    - hydrophobic surfaces is estimated at 0.03 kcal/mol/Å^2
    - Ion interactions and salt bridges much stronger
    - Complementary polar surfaces is estimated at 0.1 kcal/mol/Å^2

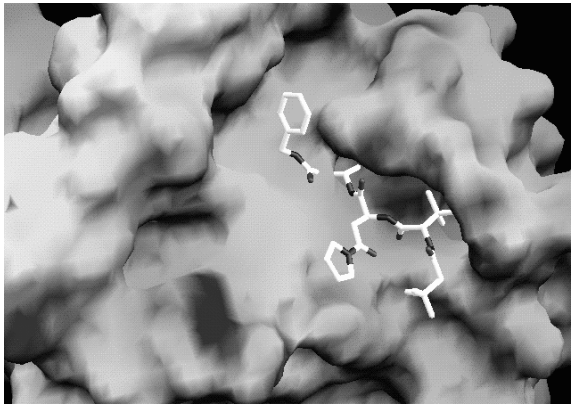


- A drug with K_i = 10nM therefore needs to bury 370 Å^2 of hydrophobic surface area.
- Every 46Å^2 of buried hydrophobic surface area buys 10-fold increase in potency
  - Maximial affinity per atom –1.5 kcal/mol

A ‘beautiful’ serine protease site:
Thrombin
(potent non-covalent Ro5 compounds)



An ‘ugly’ serine protease site:
CMV protease.
Larger than Ro5 or reactive ‘warheads’ required for potency



Methods

Inpharmatica/Pfizer Collaboration

- Database of drug targets
- Database of targets with leads
- Protein structure-based analysis
  - calculated from the protein structures including volume, depth, curvature, accessibility, hydrophobic surface area and polar surface area
  - This method has a demonstrated an 91% success rate when predicting druggability on the protein oral drug targets
- Sequence homology
  - Homology bases on BLAST cut off of 30% sequence identity and E-value less than or equal to 10^{-5}
- Feature-based Bayesian
  - Bayesian druggability model of known drug targets based on 100 protein properties and feature
- *Work is on-going and expected to be published early 2006*

Pfizer

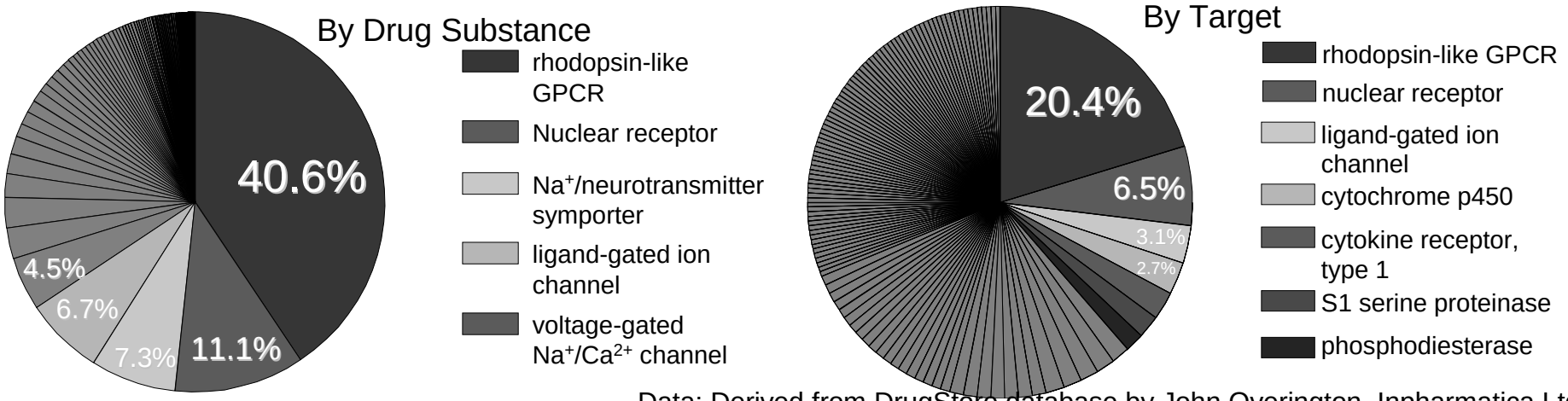
inpharmatica

Molecular Drug Targets

- Orange book, 2005
  - 26,000 drugs products which reduces to 1783 unique new molecular entities
  - of which 1415 are small molecule chemical entities
  - 180 are biological therapeutics
  - 18 of which are antibodies

Class of Drug Target	No. Molecular Targets
Targets of approved NMEs (Human and anti-infectives)	301
Human Targets of approved NMEs	238
Human Targets of approved NCEs	170
Targets of approved biologicals	59
Human Targets of approved antibodies	15

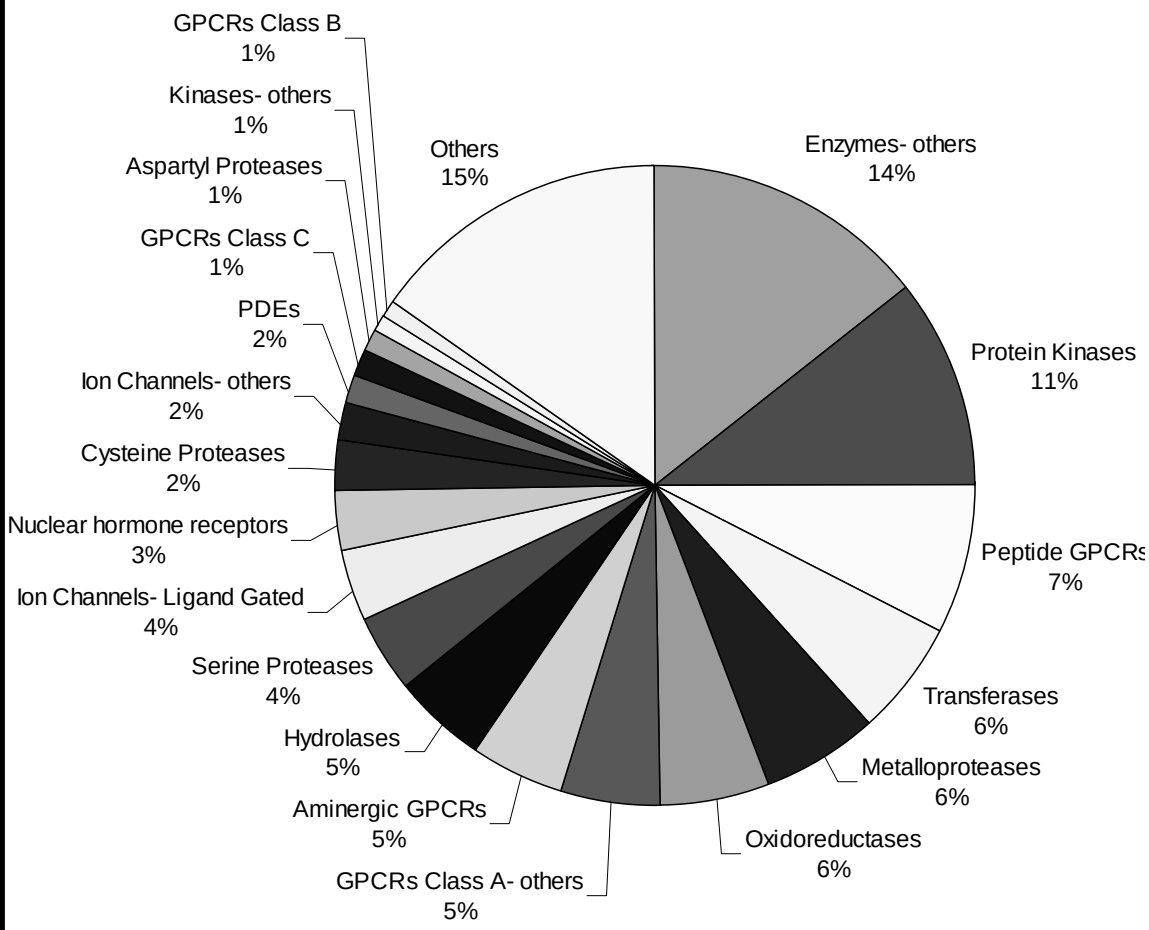
Molecular Targets of Current Drugs



Data: Derived from DrugStore database by John Overington, Inpharmatica Ltd

Targets with Leads

Gene Family	Non-redundant human targets <10uM	Ro5 Non-redundant human Targets <10uM
Aminergic GPCRs	34	34
Aspartyl Proteases	7	3
Cysteine Proteases	16	14
Enzymes- others	102	81
GPCRs Class A- others	35	30
GPCRs Class B	5	2
GPCRs Class C	10	10
Hydrolases	34	28
Ion Channels- Ligand Gated	26	20
Ion Channels- others	14	12
Kinases- others	7	6
Metalloproteases	41	39
Nuclear hormone receptors	22	19
Others	108	79
Oxidoreductases	39	37
PDEs	11	11
Peptide GPCRs	52	42
Protein Kinases	75	66
Serine Proteases	27	24
Transferases	42	30
Total	707	587



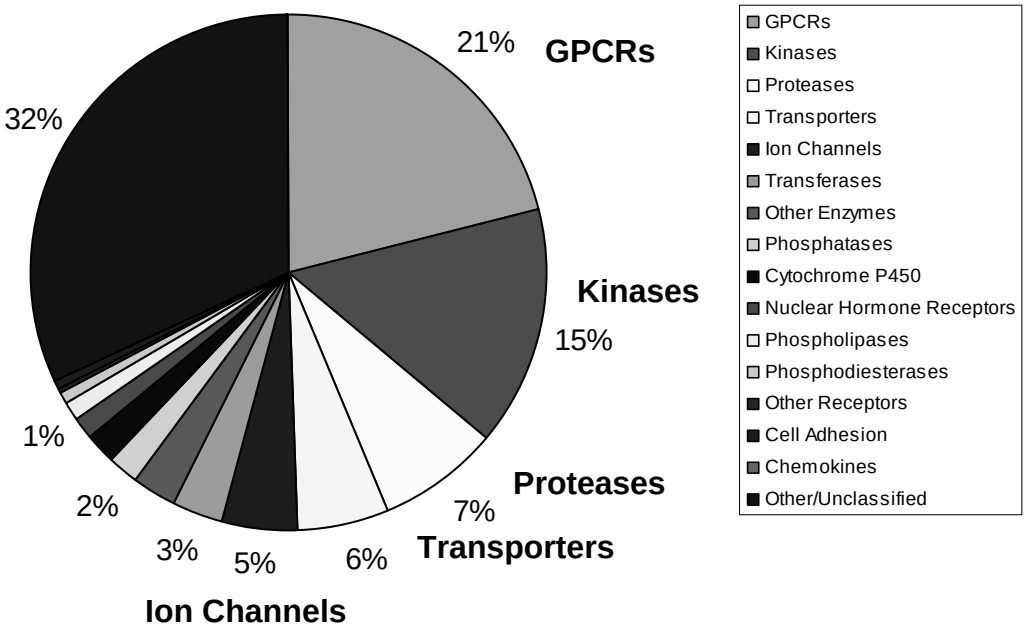
Gene Family distribution of non-redundant human proteins with small molecule chemical leads with binding affinities <10uM. Data derived from analysis of 25 years of published Med Chem data

G. Paolini

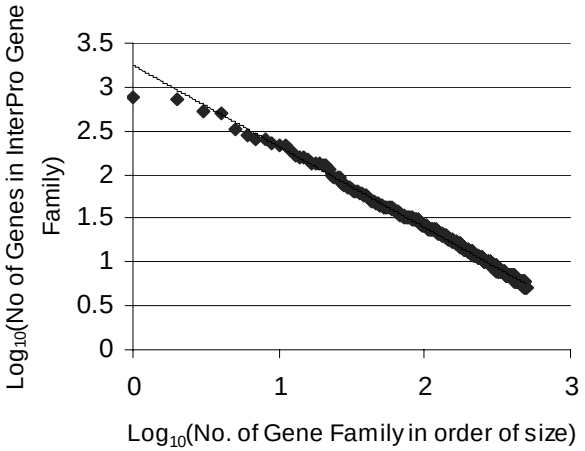
Druggable Genome Predictions

Druggability Prediction Method	No. Molecular Targets
Targets of approved NCEs	170
Sequence homology to NCE drug targets	945
Targets of chemical leads with activities (binding affinities) below 10uM	707
Targets of Ro5 chemical leads with activities (binding affinities <= 10uM)	587
Sequence homology to targets with chemical leads*	2921
Feature-based druggability sequence probability prediction	2325
Structured-based prediction	427
Sequence homology to proteins predicted druggable by structure-based method	3541
Predicted Druggbable Genome (high confidence)	**3505**
Human Genome	23000

Gene family distributions of predicted druggable genome



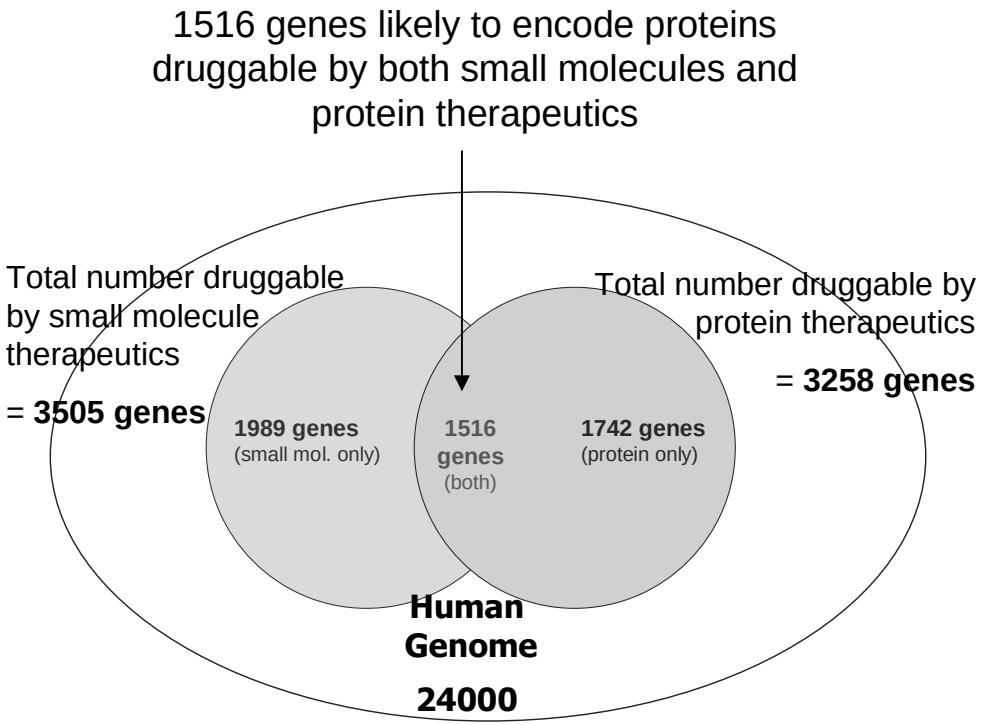
Druggable genome estimates increase slowly due to power law nature of gene family populations



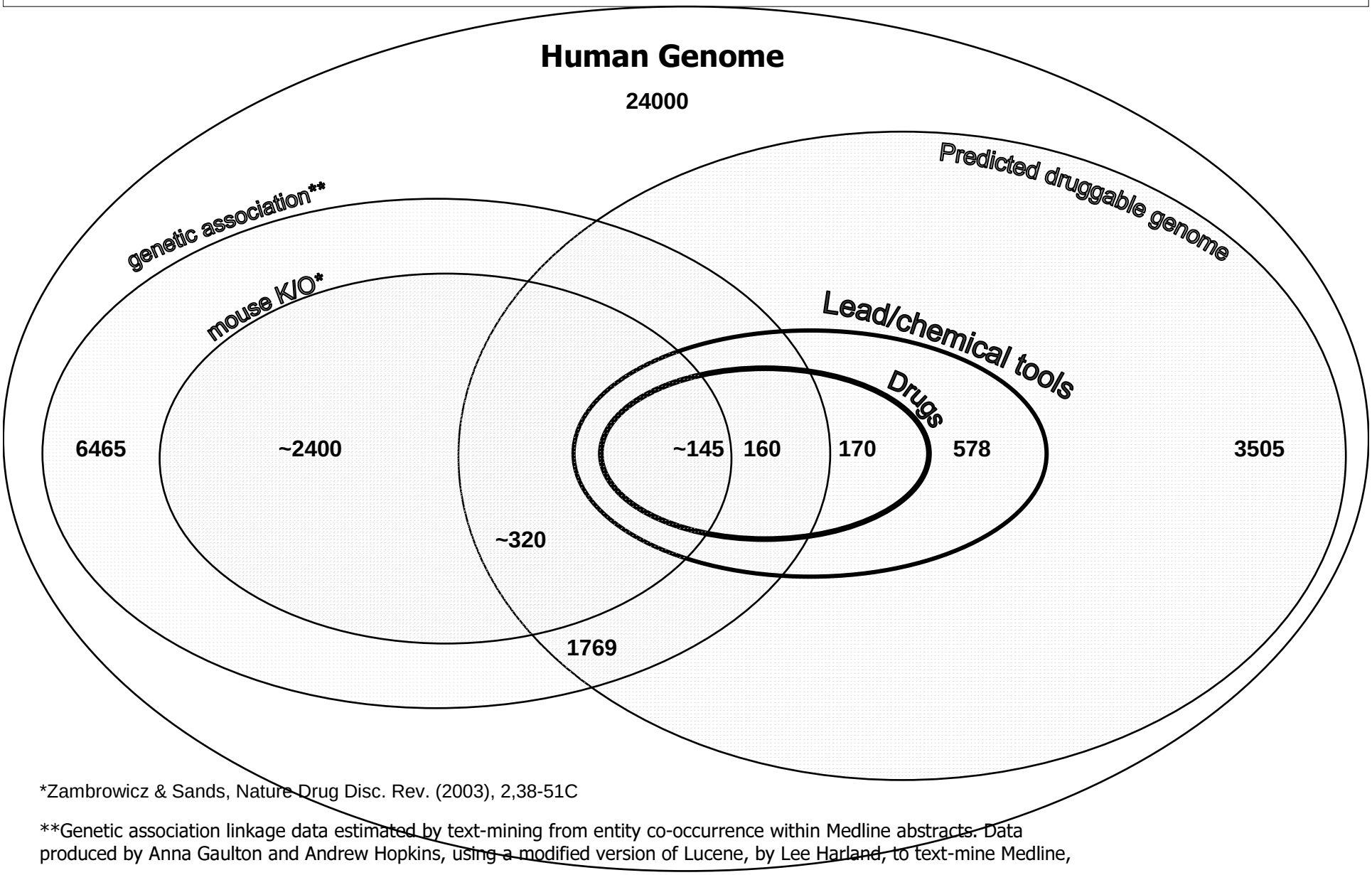
Genome data calculated by Bissan Al-Lazikani & John Overington, Inpharmatica, 2005

Protein Therapeutic Accessible Genome

Druggability Prediction Method	No. of Molecular targets
Targets of approved antibodies	15
Targets of approved biologicals	59
Secreted protein (high confidence)	1384
Secreted proteins (low confidence)	6560
Transmembrane predictions (high confidence)	973
Transmembrane predictions (low confidence)	1407
Unique, combined transmembrane and secreted predictions (high confidence)	2287
Feature-based biological target sequence probability prediction	1637
Total unique genes predicted to be accessible via protein therapeutics	**3258**

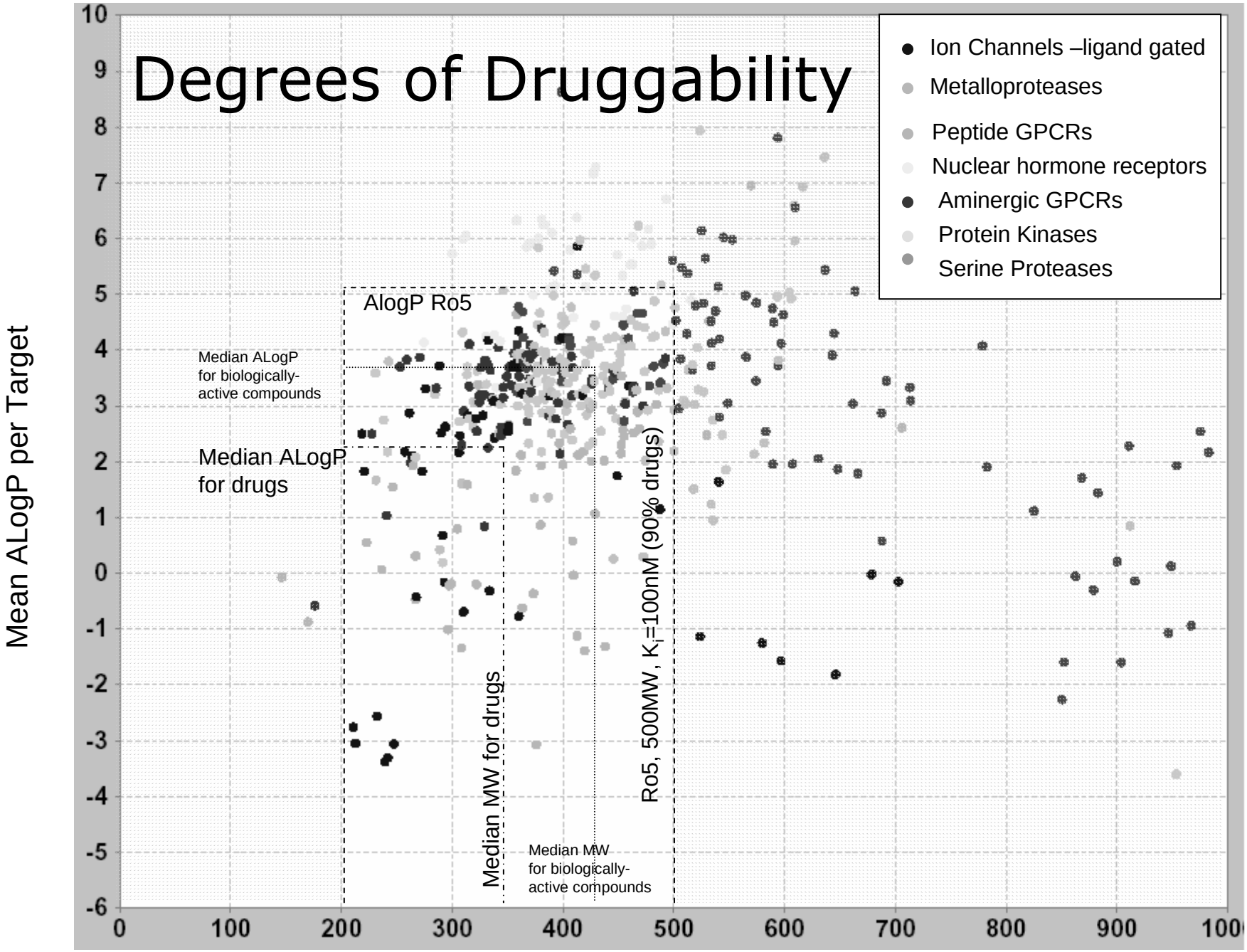


Future Drug Target Space



*Zambrowicz & Sands, Nature Drug Disc. Rev. (2003), 2,38-51C

**Genetic association linkage data estimated by text-mining from entity co-occurrence within Medline abstracts. Data produced by Anna Gaulton and Andrew Hopkins, using a modified version of Lucene, by Lee Harland, to text-mine Medline,


Degrees of Druggability
Ion Channels –ligand gated
Metalloproteases
Peptide GPCRs
Nuclear hormone receptors
Aminergic GPCRs
Protein Kinases
Serine Proteases
AlogP Ro5
Median ALogP for biologically-active compounds
Median ALogP for drugs
Median MW for drugs
Median MW for biologically-active compounds
Ro5, 500MW, K_i=100nM (90% drugs)
Mean ALogP per Target
Mean MW per target (Da) for all active <10μM

Understanding technological limits to spur innovation

- We are beginning to quantify the the technical limits of different drug technologies
    - E.g. small molecules vs antibodies vs rProteins
- We are beginning to understand the relationship between molecular target and physico-chemical properties of drugs in terms of probability
- The Druggable Genome concept is designed to focus limited resources on projects with highest chance of success.

Lessons for Innovation from other industries

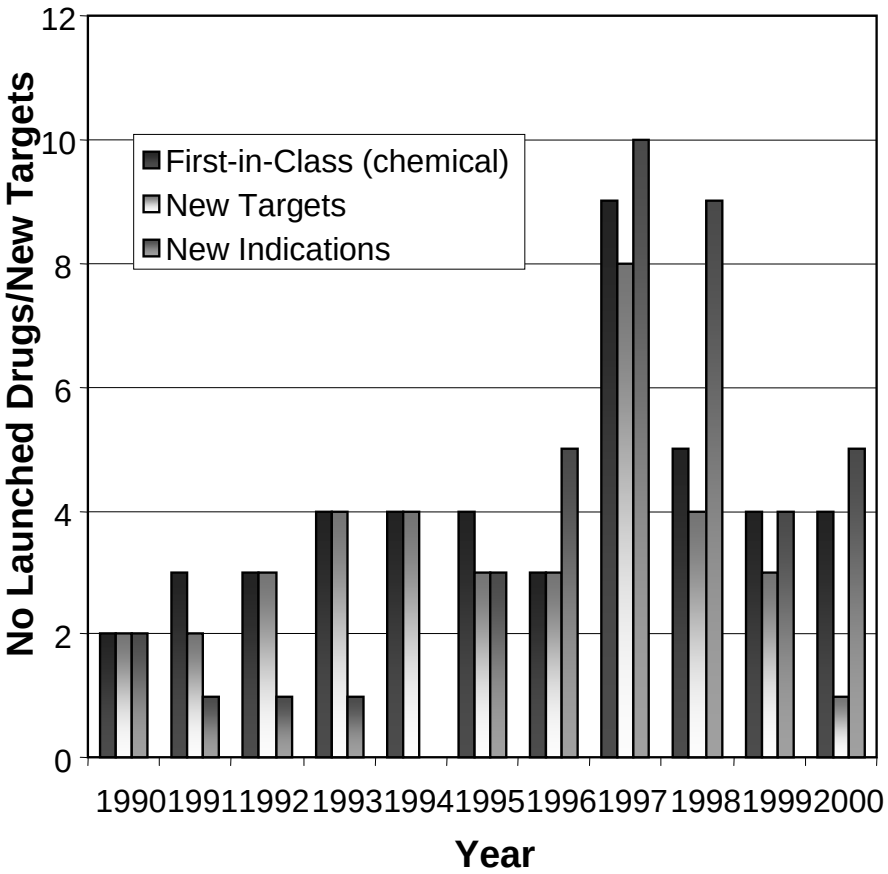
"Not only are the market applications for disruptive technologies *unknown* at the time of their development, they are *unknowable*."

– Clayton Christensen, *The Innovator's Dilemma*

- *Markets for disruptive technologies are discovered together in a dialogue between inventors and users*
  - Application of silidenafil to erectile dystfunction
- Plans for disruptive innovation must be for *learning* and *discovery* rather than execution
  - Discovery of best application of first-in-class drugs with novel mechanisms challenges market-lead TA strategies

Drugs don’t always fall into neat Therapeutic Areas

Drugs may exhibit both segmented efficacy in the target diseases *and* unexpected alternative indications



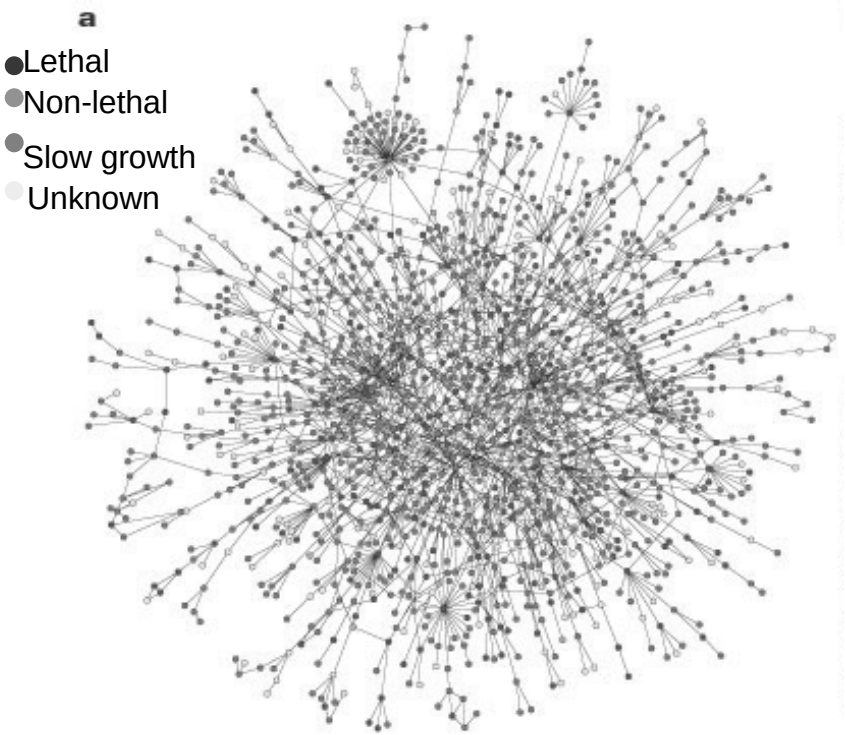
Data: Derived from Drug News Perspect, Prous Science
Gelijns *et al* New Engl J Med 339 (10) 693-8
Pritchard *et al* ‘Capturing the unexpected benefits of medical research’, OHE 2001

- 40% Sales for Alternative Indications
  - 40% of revenues of the 1993 Top 20 blockbusters came from secondary indications
  - 1999 Top 40 products
    - 62% revenues for original indication
    - 25% revenues for secondary indications
    - 13% unknown (probably 2ndy indications)
- A deeper ontological problem for drug discovery?
  - Modern basis of understanding the classification of diseases (nosology) is 300 years old
  - Molecular and etiology-based nosology is key reducing attrition due to lack of efficacy

Redundancy and Efficacy

Yeast protein-protein interaction network

Barbási *et al.* Nature, (2000), 406, 378



- Experimental evidence from phenotype observations from large-scale gene deletion studies in several model organism have shown the biological systems to be remarkably resilient to attack and perturbation
  - Biological systems can often find alternative compensatory signalling routes to bypass the inhibition of individual nodes
- The scale-free nature of biological networks are inherently resistant to random attacks predicts randomly removing (inhibiting) most targets has little effect on system
- Synthetic lethality: mutation of two gene alone leaves a cell viable but simultaneous deletions leads to death

- Re-evaluation of the role of polyphramacology:
  - Combination therapies and promiscuous drugs
- *End of the one target, one drug paradigm?*

Conclusions

- ‘High value real estate’
  - Integrating genomic data with chemical information and protein structure analysis enables the immediately identify which targets have the highest probability of success, with current technologies - in order to increase drug discovery productivity
- Learning Strategies
  - Pre-clinical & clinical discovery and learning strategies for innovating medical application of experimental drugs, with novel mechanisms, is a challenge to the dominant TA/market-lead thinking in the industry
- Polypharmacology
  - Network biology concept is providing a basis challenging the ‘one target, one drug’ paradigm
  - Combinatorial explosion of limited druggable palette

Acknowledgements

John Overington†
Bissan Al-Lazikani†
Anna Gaulton
Jerry Lanfear
Gaia Paolini
Jonathan Mason
Colin Groom‡

†Collaborators at Inpharmatica, London, UK
‡Currently at UCB, Cambridge UK

Email: andrew.hopkins@pfizer.com