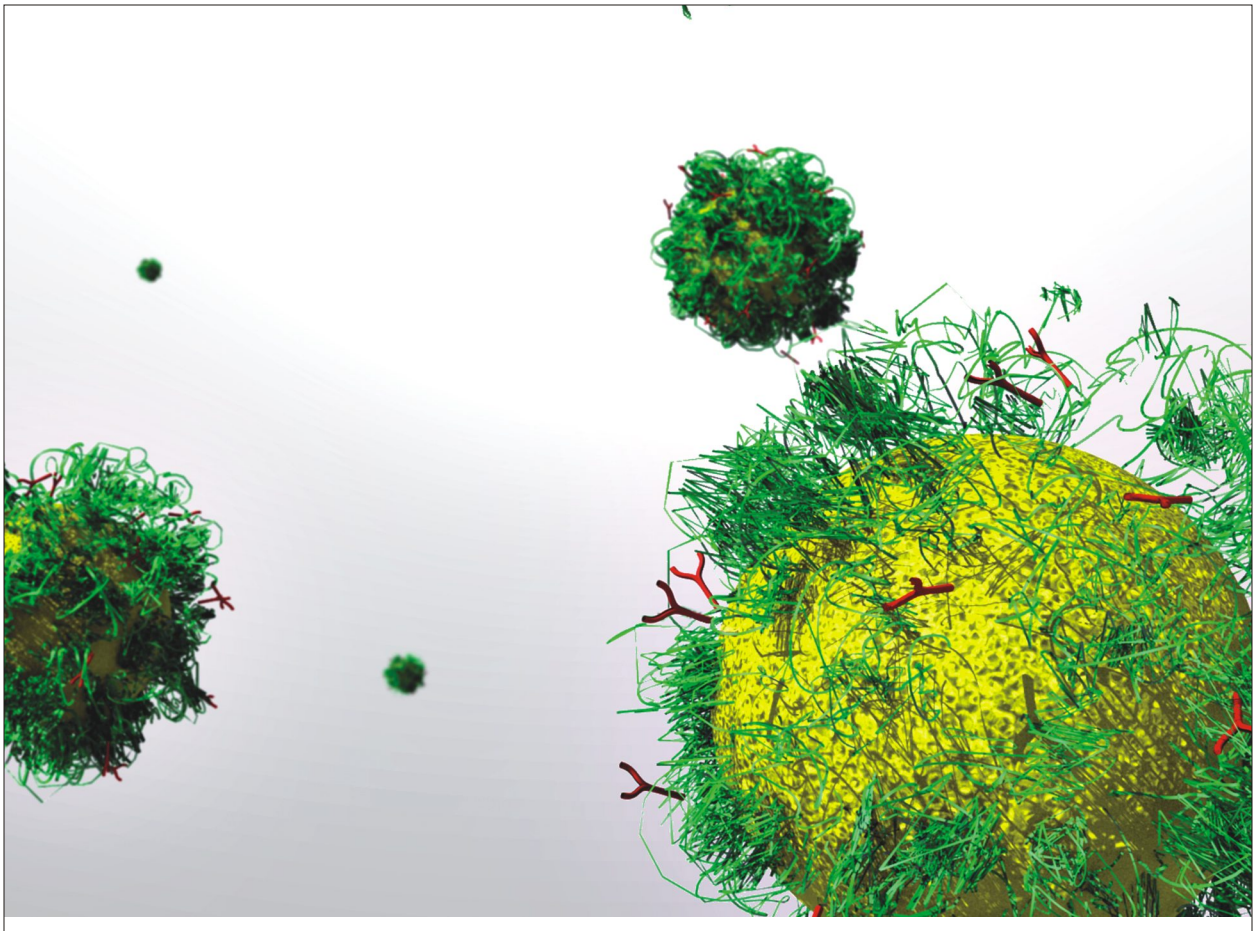

American Journal of
Drug Discovery and Development

ISSN 2150-427X

Academic
Journals Inc.

www.academicjournals.com

American Journal of Drug Discovery and Development 2 (4): 184-193, 2012
ISSN 2150-427x / DOI: 10.3923/ajdd.2012.184.193

Selective Constraint: A Hallmark of Genes Successfully Targeted for Pharmaceutical Development

[1]Qing Zhang, [2]Ada Solidar, [1]Nicholas J. Murgolo, [1]Wynand Alkema, [1]Wei Ding, [1]Peter M. Groenen, [1]Jonathan R. Greene, [3]Eric L. Gustafson, [1]Jan Klomp, [1]Ellie D. Norris, [1]Ping Qiu and [4]Gerald J. Wyckoff

[1]Global Molecular Design and Informatics, Merck Research Laboratories, USA
[2]VaSSA Informatics, USA
[3]Functional Genomics, MRL, USA
[4]UMKC SBS, Division of Molecular Biology and Biochemistry, Kansas City, Mo, USA

Corresponding Author: Gerald J. Wyckoff, UMKC SBS, Division of Molecular Biology and Biochemistry, Kansas City, MO, USA

ABSTRACT

We examined the synonymous vs. nonsynonymous substitution rate ratios (Ka/Ks, aka evolutionary rate) between human and chimpanzee for 166 successful drug target genes and compared them with a larger (10,298) set of genes representative of average human-chimpanzee evolutionary rates. We found that evolutionary rates differ significantly between successfully marketed drug targets and the broader set of genes ($p<0.005$ by ANOVA). Evolutionary rates were lower for successfully marketed drug targets versus non-target genes (0.311 versus 0.497). This rate discrepancy demonstrates that more conserved genes, even within protein families such as GPCRs (successful target GPCRs 0.391 versus non-target GPCRs 0.855) and protein kinases (0.131 versus 0.337), are better targets for traditional small molecule drug development than less strongly constrained genes. Evolutionary rate, therefore, is a factor that could be taken into account when selecting candidate target genes for drug discovery, in addition to the biochemical properties of the proteins these genes encode. We suggest therefore, that links be established between identified disease-causal or -associated genes and genes that are suitable targets for traditional small molecule pharmaceutical development.

Key words: Evolution, natural selection, drug development

INTRODUCTION

Currently, the pre-clinical phase of drug development-including target identification and validation, assay development, compound screening, transition from hit to lead, optimization of leads and finally, the development of those leads-takes an average of 4-6 years and costs, capitalized, over $30M per annum (Di Masi *et al*., 2003; Adams and Brantner, 2010). The identification of valid gene targets for therapeutic intervention early in this cycle, therefore, has long-term importance both to the public health and the economic strength of the pharmaceutical industry. With the rise of target-based drug discovery and *in silico* methods of screening, the process of target selection has become more tractable; for example, Bakheet and Doig (Bakheet and

Doig, 2009). Identified sets of desirable properties including high hydrophobicity, long length and presence of SignalP motif and Zhu (Zhu *et al.*, 2009) took into consideration sequence, structure, physicochemical characteristics and systems profiles. These studies are excellent demonstration of incorporation of protein biochemical property into drug discovery. The explosion of small molecule sources (Ramasamy *et al.*, 2011; El-Said and Al-Barak, 2011; Kayode and Kayode, 2011; Meena *et al.*, 2010; Gantait *et al.*, 2010; Arkin and Wells, 2004; Dancik *et al.*, 2010) and delivery systems (Akiyama *et al.*, 2000; Harisa *et al.*, 2011; Eleazu *et al.*, 2011; Balamuralidhara *et al.*, 2011) means that target selection will be highly important for most screening methodologies.

While the concept of examining the evolutionary rates of candidate genes is not new (Ma and Wang, 2009; Durand *et al.*, 2008; Bradford *et al.*, 2006), to date there has been no consensus on what the best features to examine are from a pharmaceutical perspective (Zhu *et al.*, 2009). Is success more likely to come from targeting drugs at a gene that is evolving more rapidly, or evolving more conservatively? Alternatively, is there no correlation at all? Previous studies have found that disease-causal genes are likely to be more rapidly evolving (Longman-Jacobsen *et al.*, 2003; Dawkins *et al.*, 1999; Fay and Wittkopp, 2008; Ma and Wang, 2009; Durand *et al.*, 2008; Blekhman *et al.*, 2008; Holbrook and Sanseau, 2007) than the general rate of primate gene evolution. Taken with previous genomic studies (Winter *et al.*, 2004; Bradford *et al.*, 2006; Smith and Eyre-Walker, 2003; Vamathevan *et al.*, 2008), our current study strongly suggests that disease genes and successful drug target genes are often different, incongruent sets. In this article, we also extend our analysis to include successful drug targets.

We examined the properties of drug target genes that have yielded successful drugs on the market. We asked if evolutionary rate is a viable trait to examine when attempting to narrow the analysis of drug targets from a broad list of potential drug target candidates.

MATERIALS AND METHODS

Orthologous gene pair acquisition and analysis: Data was acquired, in part, from the SPEED database (Vallender *et al.*, 2006). When data from SPEED was not available or alignments were redone, the following procedure was followed. Datasets used were obtained from Ensembl. For purposes of this paper, we focused on human and chimp sequences (Fernandez-Banet *et al.*, 2007).

When not already available, human-mouse orthologs were gathered from synteny maps and sequence homology (Eyre *et al.*, 2007; Wheeler *et al.*, 2006; Shah *et al.*, 2005) and the sequences for each gene were recovered from RefSeq. Next, mouse-rat orthologs were identified by Blast (Altschul *et al.*, 1997; Altschul *et al.*, 2005). Rat orthologs to mouse genes were identified as those that were reciprocal best hits when compared with a mouse data file (from RefSeq) (Wheeler *et al.*, 2006) using nucleotide Blast, with an e-value cut off of 1 x 10-50. Coding region sequences annotated from RefSeq were used to speed up the Blast search. In many cases, other supporting information was available to corroborate orthology (e.g., gene name annotation or orthology reported in other literature). From SPEED, a total of 13,506 groups containing orthologs in primate species, with 10,929 of these also containing mouse and rat orthologs. Alignments were re-checked utilizing CLUSTAL on translated sequences frame-aligned with nucleotide sequences, preserving gaps in the amino acid alignment (Thompson *et al.*, 2002; Chenna *et al.*, 2003). After considering only our confirmed ortholog alignments and removing cases where translation

frame was uncertain, a total of 10,465 genes were available for analysis. Preliminary data analysis was performed using BioVassa, a statistical tool created by Vassa Informatics to assess information content (Moldover *et al*., 2009).

We determined the synonymous substitution rate (Ks) and the nonsynonymous substitution rate (Ka) using the (Li, 1993) method as employed in the Diverge program from Accelrys (San Diego) This ratio is commonly employed as a measure of constraint on gene evolution. Alternative methodologies for determining Ka/Ks were utilized and produced similar results throughout.

Target gene identification and functional sorting: Lists of targets were gathered by utilizing literature searches and lists of known drug targets gathered from public data sources (Bjarnadottir *et al*., 2006; Vassilatis *et al*., 2003; Kostich *et al*., 2002; Manning *et al*., 2002). Successful targets refer to genes utilized for drug development that lead to successfully marketed drugs and Research targets were identified as those candidate genes which are within current research pipelines but for which no drug has been marketed. Genes were identified as GPCRs (Bjarnadottir *et al*., 2006; Vassilatis *et al*., 2003; Takeda *et al*., 2002) or Protein Kinases (Kostich *et al*., 2002; Manning *et al*., 2002) based upon a review of their function from literature, Gene Ontology (GO) terms (Ashburner *et al*., 2000) (searched August, 2009) and Ingenuity Pathway Analysis (Leschly and Basset, 2009).

Statistical evaluation: Statistical analysis was performed utilizing Microsoft Excel (11.6560.6568) to examine variance and covariance of evolutionary rates within and between gene sets. Genes with missing data were excluded from these analyses.

RESULTS

An analysis of successful drug targets versus a representative set of genes showed a highly significant difference in the evolutionary rates of these gene sets (p-value, 0.003, Table 1). Raw data tables are available upon request.

This is mirrored when successful targets which are GPCRs or Kinases are examined (p-value, 0.046 and 0.047, respectively; Table 2).

Table 1: Comparison of primate ortholog in successful drug target genes vs. non-target or all genes

Group	Count	Average human-chimpanzee Ka/Ks	p-value vs. successful targets
Successful targets	166	0.311	na
Research targets	501	0.399	0.04
Non-targets	9631	0.497	0.003
All genes	10298	0.489	0.003

Table 2: Comparison of primate ortholog in successful GPCR and protein kinase drug target vs. non-targets or all members of the family

Protein family	Group	Count	Average human-chimpanzee Ka/Ks	p-value vs. successful targets
GPCR	Successful targets	36	0.391	na
	Non-targets	316	0.855	0.018
	All genes	420	0.748	0.046
Protein kinase	Successful targets	19	0.131	na
	Non-targets	235	0.337	0.028
	All genes	311	0.309	0.047

Table 3: Human primate orthologs of disease genes

	Count	Average human-chimpanzee Ka/Ks	p-value vs. disease genes
Disease genes	104	0.473	na
OMIM genes	71	0.424	
GWAS genes	34	0.568	
All genes	10298	0.489	0.834

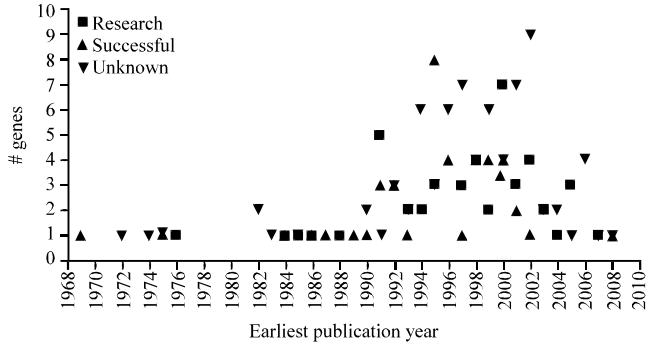

Fig. 1: Gene and publication dates, three categories of genes were sampled, successful, research and unknown, p value among all pairs were >0.05. This chart shows the number of genes in the successful and research categories and the earliest publication date relating to that gene or protein

To compare the human primate conservation of disease genes with drug target genes, we collected 71 confirmed disease-causal genes from Online Mendelian Inheritance in Man **(OMIM, 2006)** **(OMIM) and 34 disease-associated genes from Genome Wide Association Studies** (Altshuler *et al*., 2008) (GWAS) (with coding-nonsynonymous SNP and a p-value <= 1E-3). In contrast to the successful drug targets, the human primate Ka/Ks of disease genes does not show statistically significant difference when compared to all genes (Table 3).

We further considered whether or not there was ascertainment bias caused by publication date. The hypothesis here was that more conserved genes may have been under examination longer due to genetic manipulations of lower organisms that became available at an earlier date. These genes may be more likely to be successful drug targets, due solely to the amount of time they have been studied. Our analysis showed that there was no correlation between date of publication for a gene and the presence of a gene in the successful drug targets set (Fig. 1). When 168 genes were sampled randomly, both research (n = 47) and successful (n = 43) gene targets showed a similar pattern of publication dates; this is mirrored when a set of 78 non-targets was examined (p-value >0.05 for all comparison).

DISCUSSION

Disease genes are significantly more rapidly evolving than non-disease genes, particularly when tissue specificity is considered (Smith and Eyre-Walker, 2003; Blekhman *et al*., 2008; Holbrook and Sanseau, 2007; Vamathevan *et al*., 2008). Our analysis shows that genes which have successfully been developed into drug targets are, however, generally more conserved. Discovery

of genes whose proteins are targets for successful drug development is one of the lynchpins of pharmaceutical research. Mirroring this, the goal of many human disease-mapping studies is the identification and analysis of genes where variant alleles cause protein products causal or contributory to human disease conditions. There is, in many cases, an inherent assumption that discovering a human disease gene will directly lead to a candidate gene for pharmaceutical research. Our results, taken in the context of previous literature, suggest otherwise. Our analysis strongly suggests that the process of treating a disease with a drug is very distinct from the process of identifying causal genes. That is, while variations in rapidly-evolving genes may tend to cause disease, it is more clock-like, slowly evolving genes that are better candidates as drug targets. There are a number of reasons why this may be so.

Tissue specificity is a likely culprit. Several studies found that genes expressed in specific tissues are more likely to be rapidly evolving than those expressed in multiple tissues (Duret *et al*., 2002; Duret and Mouchiroud, 2000; Duret and Mouchiroud, 1999; Duret *et al*., 1995; Duret *et al*., 1994). However, housekeeping genes, which are not often found to be involved in human diseases, are typically highly conservative; likely, diseases caused by these conserved genes are so, deleterious that they are incompatible with development and viability (Santibanez Koref *et al*., 2003). However, conservative genes appear to be better targets for drug development, given our results. This brings up two related issues: first, that conservative genes are likely to have broader tissue expression and secondly that they may be involved with basal functionality (Duret and Mouchiroud, 2000). More conserved genes may represent suitable drug development targets as behavior between species may be more predictable. Agents may bind species homologs with similar affinity and in similar orientations, making transition from pre-clinical primate model data to clinical human data more reliable.

Our analysis specifically explored the possibility that specific functions may be contributing to the observed bias in evolutionary rate in drug target genes. Our analysis examines, specifically, GPCR and Kinase proteins. In this case, the observed difference between successful targets and all proteins remained in these subclasses. This suggests that specific functionality is not the likely explanation for the observed evolutionary rate bias.

We also specifically were concerned with ascertainment bias due to length of study of specific proteins. Our analysis ruled out this bias, suggesting that "first in, first out" drug development is not necessarily the norm; that is, that those proteins that have been studied the longest are not likely to yield drugs more rapidly than novel targets. Fresh drug targets, chosen rationally based on conservation, may be as likely to eventually generate a drug as longer studied proteins. This approach may be extensible to human disease gene datasets (Bakheet and Doig, 2009) as well as databases of other species, such as human pathogens (Shakyawar *et al*., 2011) and even plant pathogens (Kizil *et al*., 2005).

This finding explains, in part, why the large number of human disease genes identified has not led to a similar explosion in successful drug targets. It also suggests a rational approach to determining new targets for drug development following the identification of a gene involved with human disease. Instead of studying the human disease gene and suspecting it will lead to a drug, these results suggest that analyzing for example the pathway that the gene is in and picking a conservative protein within that pathway may be a better approach to drug design. For example, polymorphism of endothelin 1, a potent vasoconstrictor, has been associated with hypertension (Jin *et al*., 2003). When the human-chimpanzee Ka/Ks value was overlayed on endothelin 1

pathway in Ingenuity Pathway Analysis (http://www.ingenuity.com) (Leschly and Basset, 2009) (Fig. 2 and Table 4), it was noted that its receptor, a GPCR named endothelin receptor type B (EDNRB or ETB), has a much smaller human-chimpanzee Ka/Ks value (0.307) than endothelin

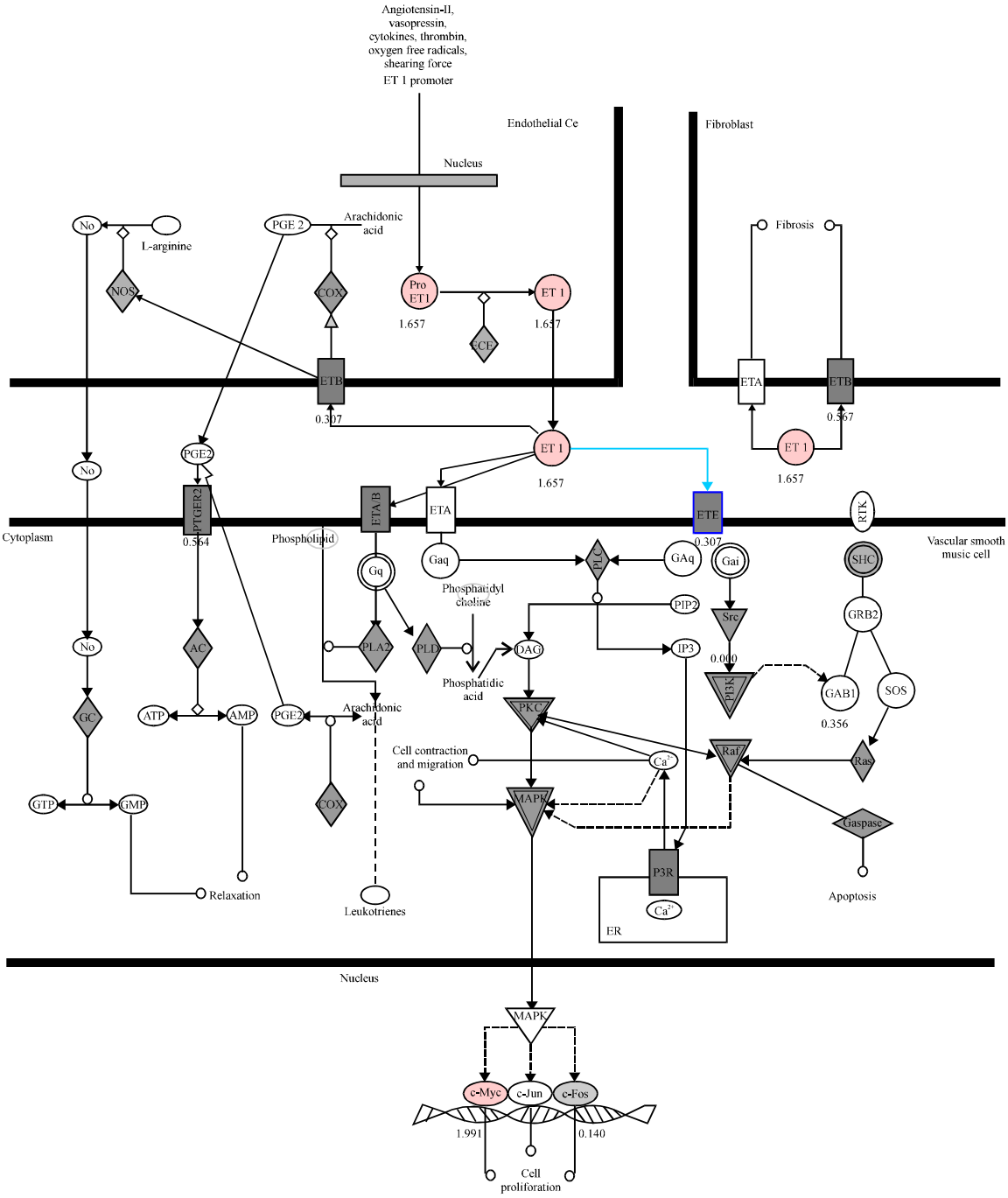

Fig. 2: The endothelin-1 signaling pathway, The Ka/Ks values (in small numbers) are high for ET1 (Endothelin-1), but low for EDNRB (labelled as ETE). Bosentan, a recently licensed drug targeted against hypertension, targets EDNRB

Table 4: Human primate ortholog in endothelin 1 pathway

Family	Drugs	Entrez gene name	Symbol	Location	Human chimpanzee Ka/Ks
Kinase	Dasatinib, AZM-475271, AZD 0530	V-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	Src	Cytoplasm	0.0
Transcription regulator		FBJ murine osteosarcoma viral oncogene homolog	c-Fos	Nucleus	0.139537
G-protein coupled receptor	Bosentan, sitaxsentan, atrasentan	Endothelin receptor type B membrane	ETB	Plasma	0.1307144
Other		GRB2-associated binding protein 1	GAB1	Cytoplasm	0.356205
G-protein, coupled receptor	Misoprostol, prostaglandin E2, prostaglandin E1, CP 533536, diclofenac/misoprostol	Prostaglandin E receptor 2, (subtype EP2), 53 kDa	PTGER2	Plasma membrane	0.563919
Other		Endothelin 1	ET1	Extracellular space	1.65737
Transcription regulator	v-myc myelocytomatosis viral oncogene homolog (avian)	c-Myc	Nucleus	1.99129	

1 (1.657), making it a potentially more successful drug target than endothelin 1. Indeed, consistent with our hypothesis, a known inhibitor of EDNRB, Bosentan (Tracleer) has been approved for pulmonary hypertension (Krum *et al*., 1998).

REFERENCES

Adams, C.P. and V.V. Brantner, 2010. Spending on new drug development. Heath Econ., 19: 130-141.

Akiyama, K., H. Watanabe, S. Tsukada and H. Sasai, 2000. A novel method for constructing gene-targeting vectors. Nucleic Acids Res., 28: E77-E77.

Altschul, S.F., J.C. Wootton, E.M. Gertz, R. Agarwala, A. Morgulis, A.A. Schaffer and Y.K. Yu, 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J., 272: 5101-5109.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl. Acids Res., 25: 3389-3402.

Altshuler, D., M.J. Daly and E.S. Lander, 2008. Genetic mapping in human disease. Science, 322: 881-888.

Arkin, M.R. and J.A. Wells, 2004. Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream. Nat. Rev. Drug Discov., 3: 301-317.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein and D. Butler *et al*., 2000. Gene Ontology: tool for the unification of biology: The gene ontology consortium. Nat. Genet., 25: 25-29.

Bakheet, T.M. and A.J. Doig, 2009. Properties and identification of human protein drug targets. Bioinformatics, 25: 451-457.

Balamuralidhara, V., T.M. Pramodkumar, N. Srujana, M.P. Venkatesh, N.V. Gupta, K.L. Krishna and H.V. Gangadharappa, 2011. pH sensitive drug delivery systems: A review. Am. J. Drug Discovery Dev., 1: 24-48.

Bjarnadottir, T.K., D.E. Gloriam, S.H. Hellstrand, H. Kristiansson, R. Fredriksson and H.B. Schioth, 2006. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. Genomics, 88: 263-273.

Blekhman, R., O. Man, L. Herrmann, A.R. Boyko and V. Indap *et al*., 2008. Natural selection on genes that underlie human disease susceptibility. Curr. Biol., 18: 883-889.

Bradford, J.R., C.J. Needham, A.J. Bulpitt and D.R. Westhead, 2006. Insights into protein-protein interfaces using a Bayesian network prediction method. J. Mol. Biol., 362: 365-386.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.J. Higgins and J.D. Thompson, 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res., 31: 3497-3500.

Dancik, V., K.P. Seiler, D.W. Young, S.L. Schreiber and P.A. Clemons, 2010. Distinct biological network properties between the targets of natural products and disease genes. J. Am. Chem. Soc., 132: 9259-9261.

Dawkins, R., C. Leelayuwat, S. Gaudieri, G. Tay, J. Hui, S. Cattley, P. Martinez and J. Kulski, 1999. Genomics of the major histocompatibility complex: Haplotypes, duplication, retroviruses and disease. Immunol. Rev., 167: 275-304.

Di Masi, J.A., R.W. Hansen and H.G. Grabowski, 2003. The price of innovation, new estimates of drug development costs. J. Health Econom., 22: 151-185.

Durand, P.M., K. Naidoo and T.L. Coetzer, 2008. Evolutionary patterning: A novel approach to the identification of potential drug target sites in *Plasmodium falciparum*. PloS One, Vol. 3. 10.1371/journal.pone.0003685.

Duret, L., D. Mouchiroud and M. Gouy, 1994. HOVERGEN: A database of homologous vertebrate genes. Nucleic Acids Res., 22: 2360-2665.

Duret, L., D. Mouchiroud and C. Gautier, 1995. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J. Mol. Evol., 40: 308-317.

Duret, L. and D. Mouchiroud, 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila* and *Arabidopsis*. Proc. Natl. Acad. Sci., 96: 4482-4487.

Duret, L. and D. Mouchiroud, 2000. Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. Mol. Biol. Evol., 17: 68-74.

Duret, L., M. Semon, G. Piganeau, D. Mouchiroud and N. Galtier, 2002. Vanishing GC-rich isochores in mammalian genomes. Genetics, 162: 1837-1847.

El-Said, S.M. and A.S. Al-Barak, 2011. Extraction of insulin like compounds from bitter melon plants. Am. J. Drug Discovery Dev., 1: 1-7.

Eleazu, C.O., J.U. Amajor, A.I. Ikpeama and E. Awa, 2011. Studies on the nutrient composition, antioxidant activities, functional properties and microbial load of the flours of 10 elite cassava (*Manihot esculenta*) varieties. Asian J. Clin. Nutr., 3: 33-39.

Eyre, T.A., M.W. Wright, M.J. Lush and E.A. Bruford, 2007. HCOP: A searchable database of human orthology predictions. Brief. Bioinform., 8: 2-5.

Fay, J.C. and P.J. Wittkopp, 2008. Evaluating the role of natural selection in the evolution of gene regulation. Heredity, 100: 191-199.

Fernandez-Banet, J., S. Graf, S. Haider, M. Hammond and J. Herrero *et al*., 2007. Ensembl 2007. Nucleic Acids Res., 35: D610-D177.

Gantait, S., N. Mandal and P.K. Das, 2010. An overview on *in vitro* culture of genus *Allium*. Am. J. Plant Physiol., 5: 325-337.

Harisa, G.I., M.F. Ibrahim, F.K. Alanazi and I.A. Alsarra, 2011. Application and safety of erythrocytes as a novel drug delivery system. Asian J. Biochem., 6: 309-321.

Holbrook, J.D. and P. Sanseau, 2007. Drug discovery and computational evolutionary analysis. Drug Discov. Today, 12: 826-832.

Jin, J.J., J. Nakura, Z. Wu, M. Yamamoto and M. Abe *et al*., 2003. Association of endothelin-1 gene variant with hypertension. Hypertension, 41: 163-167.

Kayode, A.A.A. and O.T. Kayode, 2011. Some medicinal values of *Telfairia occidentalis*: A review. Am. J. Biochem. Mol. Biol., 1: 30-38.

Kizil, S., F. Uyar and A. Sagir, 2005. Antibacterial activities of some essential oils against plant pathogens. Asian J. Plant Sci., 4: 225-228.

Kostich, M., J. English, V. Madison, F. Gheyas and L. Wang *et al*., 2002. Human members of the eukaryotic protein kinase family. Genome Biol.

Krum, H., R.J. Viskoper, Y. Lacourciere, M. Budde and V. Charlon, 1998. The effect of an-endothelin-receptor antagonist, bosentan, on blood pressure in patients with essential hypertension: Bosentan hypertension investigators. N. Engl. J. Med., 338: 784-790.

Leschly, J. and D. Basset, 2009. The Ingenuity Pathway Analysis Tool. Ingenuity Systems, Redwood City, CA. http://www.ingenuity.com/

Li, W.H., 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol., 36: 96-99.

Longman-Jacobsen, N., J.F. Williamson, R.L. Dawkins and S. Gaudieri, 2003. In polymorphic genomic regions indels cluster with nucleotide polymorphism: Quantum Genomics. Gene, 312: 257-261.

Ma, X. and Z. Wang, 2009. Anticancer drug discovery in the future: An evolutionary perspective. Drug Discov. Today, 14: 1136-1142.

Manning, G., D.B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, 2002. The protein kinase complement of the human genome. Science, 298: 1912-1934.

Meena, A.K., P. Bansal, S. Kumar, M.M. Rao and V.K. Garg, 2010. Estimation of heavy metals in commonly used medicinal plants: a market basket survey. Env. Monitor Assess, 170: 657-660.

Moldover, B., J. Murphy, A. Solidar and G.J. Wyckoff, 2009. The Chemvassa Analytical Tool. Vassa Informatics., Texarkana, AR. http://www.vassainformatics.com

OMIM, 2006. Online Mendelian Inheritance in Man. National Library of Medicine, Bethesda, MD USA.

Ramasamy, M.S., P. Arumugam, S. Manikandan and A. Murugan, 2011. Molecular and combinatorial array of therapeutic targets from conotoxins. Am. J. Drug Discovery Dev., 1: 49-57.

Santibanez Koref, M.F., R. Gangeswaran, I.P. Santibanez Koref, N. Shanahan and J.M. Hancock, 2003. A phylogenetic approach to assessing the significance of missense mutations in disease genes. Hum. Mutat., 22: 51-58.

Shah, S.P., Y. Huang, T. Xu, M.M. Yuen, J. Ling and B.F.F. Ouellette, 2005. Atlas- a data warehouse for integrative bioinformatics. BMC Bioinform., Vol. 6. 10.1186/1471-2105-6-34

Shakyawar, S.K., A. Goyal and V.K. Dubey, 2011. Database of *in silico* predicted potential drug target proteins in common bacterial human pathogens. Am. J. Drug Discovery Dev., 1: 70-74.

Smith, N.G. and A. Eyre-Walker, 2003. Human disease genes: Patterns and predictions. Gene, 318: 169-175.

Takeda, S., S. Kadowaki, T. Haga, H. Takaesu and S. Mitaku, 2002. Identification of G protein-coupled receptor genes from the human genome sequence. FEBS Lett., 520: 97-101.

Thompson, J.D., T.J. Gibson and D.G. Higgins, 2002. UNIT 2.3 multiple sequence alignment using clustalW and clustalX. Curr Protoc Bioinformatics, 10.1002/0471250953.bi0203s00

Vallender, E.J., J.E. Paschall, C.M. Malcom, B.T. Lahn and G.J. Wyckoff, 2006. SPEED: A molecular-evolution-based database of mammalian orthologous groups. Bioinformatics, 22: 2835-2837.

Vamathevan, J.J., S. Hasan, R.D. Emes, H. Amrine-Madsen and D. Rajagopalan *et al.*, 2008. The role of positive selection in determining the molecular cause of species differences in disease. BMC Evol. Biol., 8: 273-273.

Vassilatis, D.K., J.G. Hohmann, H. Zeng, F. Li and J.E. Ranchalis *et al.*, 2003. The G protein-coupled receptor repertoires of human and mouse. Proc. Natl. Acad. Sci. USA., 100: 4903-4908.

Wheeler, D.L., T. Barrett, D.A. Benson, S.H. Bryant and K. Canese *et al.*, 2006. Database resources of the national center for biotechnology information. Nucleic Acids Res., 34: D173-D180.

Winter, E.E., L. Goodstadt and C. P. Ponting, 2004. Elevated rates of protein secretion, evolution and disease among tissue-specific genes. Genome Res., 14: 54-61.

Zhu, F., L.Y. Han, C. Zheng, B. Xie and M.T. Tammi *et al.*, 2009. What are next generation innovative therapeutic targets? clues from genetic, structural, physicochemical and systems profiles of successful targets. J. Pharmacol Exp Ther, 330: 304-315.