Nature's contribution to today's pharmacopeia

To the Editor:

Two decades after mainstream drug discovery shifted its focus away from bioprospecting and natural products (NPs)[1], a systematic survey of new molecular entities (NMEs) approved by the US Food and Drug Administration (FDA) indicates that NPs still make a substantial contribution to today's new drugs[2,3]. For the 132 drugs approved by the FDA from 2008 to 2012, ~30% originated from natural sources. The analysis presented here suggests that the contribution of NPs to recent NME development remains robust and may be under appreciated. What's more, we identify key traits in NME leads originating from natural sources that can inform drug discovery efforts going forward.

We commence our analysis by assessing the different natural sources of today's NMEs. Using the literature as a reference[3], NMEs can be divided into three categories: NPs; NP derivatives (semisynthetic derivatives, mimetics and pharmacophore-guided synthetic molecules); and biologics derived from nonhuman sources (including peptides, proteins, antibodies and nucleic acids)[2]. Of these, by far the most common source of NMEs is NP derivatives. Overall, between ~25% and 40% of all NMEs of natural origin can be classified as NP derivatives. Specifically, 33/132 (25.0%), 47/120 (39.2%), 57/151 (37.7%) and 426/1,708 (24.9%) of NMEs approved by FDA were NP derivatives for the time periods 2012–2008, 2007–2003, 2002–1998 and pre-1998, respectively.

Notably, NPs were much more common in NMEs approved before 1998. They represented 6/132 (4.5%), 4/120 (3.3%), 6/151 (4.0%) and 226/1,078 (13.2%) of NMEs for the time periods 2012–2008, 2007–2003, 2002–1998 and pre-1998, respectively. This is perhaps not surprising; before 1998, drug makers were able to exploit the low-hanging fruit from NPs that could be coopted for NME approval. Once the treasure chest of NPs from centuries of medicine had been exploited, discovery of new NMEs from NPs became more challenging.

The final category of NMEs, those originating from biologics, represented 4/132 (3.0%), 4/120 (3.3%), 4/151 (2.6%) and 33/1,708 (1.9%) of NMEs for the same respective time periods as above (**Table 1** and **Supplementary Table 1**). Notably, the number of biologics has remained small but relatively stable over the entire timespan of this study.

Although the number of NPs and NP derivatives has steadily decreased since 1998 (63/151, 51/120 and 39/132 of NMEs in 1998–2002, 2003–2007 and 2008–2012, respectively), their share of new drugs only started to decline in 2008–2012 (41.7%, 40.4%, 42.5% and 29.5% in pre-1998, 1998–2002, 2003–2007 and 2008–2012, respectively).

One explanation for this decline in NPs and NP derivatives since 1998 is the shift since 1997 away from NPs to combinatorial libraries of synthetic chemicals[1]. Given the 10- to 15-year drug development time lag[1], it would not be surprising to see a dip in NMEs originating from these molecules. Nonetheless, NPs and NP derivatives still represent a substantial share of new NMEs and this share shows no acceleration in decline for the period 2008–2012 (**Table 1**).

It is possible that this set of NMEs from natural sources is primarily NP era leftovers that entered development before 1997. To evaluate this possibility, we surveyed the development history of 32 NPs and NP derivatives approved between 2008 and 2012 (**Supplementary Table 2**) and found 21/132 (65.6%) NMEs are post-NP era products with their development started around 1997–2000 (37.5%) and 2000–2004 (28.1%).

The development history of these 21 post-NP era NMEs further reveals what has contributed to their selection and development in competition with other leads. 10/21 (47.6%) were bioactive NPs, 2/21 (9.5%) were ligands originating from natural sources, 6/21 (28.6%) were NP derivatives and 3/21 (14.3%) were synthetic analogs of ligands originating from natural sources. At the start of their development, the targets of 5 (23.8%) of these post-NP era NME leads were not yet drugged (carfilzomib, everolimus, linagliptin, telaprevir and crofelemer target 20S proteasome, mTOR, dipeptidyl peptidase 4, hepatitis C virus NS3/4A protease, and CFTR and CaCC channel, respectively). The 23.8% rate for these post-NP era NME leads is comparable to the rate of previously undrugged targets addressed by synthetic drugs (17.9% rate)[4]. Thus, NPs and NP derivatives may still offer promise for drugs addressing new targets.

Existing NP drug scaffolds appear to be productive templates for deriving new drugs with 16/21 (76.2%) NME leads derived from 8 preexisting NP drug scaffold groups, including G protein–coupled receptor (GPCR)-binding peptide hormones (4 leads), macrolides (3 leads), nucleotides/nucleosides (3 leads), cephalosporins (2 leads), progestogens (1 lead), statins (1 lead), taxanes (1 lead) and xanthines (1 lead). The high percentage of new drugs derived from preexisting drug scaffolds is consistent with the report that drug-like bioactive compounds of specific target classes cluster in specific regions of chemical space[1]. The leads of four of the five drugs outside preexisting drug scaffold groups (carfilzomib, telaprevir, dabigatran and romidepsin) are peptides (epoxyketone oligopeptide, NS5A-5B substrate peptide, thrombin-interacting fragment of fibrinogen and depsipeptide cyclic structure, respectively). These further show the usefulness of peptides in deriving target-selective drugs against such difficult target classes as proteases[5] and transferases[6]. Also consistent with reports that most nature-derived drugs are from preexisting drug-productive species families and clusters[3], we found 18/21 (85.7%) NME leads are from preexisting drug-productive species families and 2/21 (9.5%) leads are from previously unexplored families in preexisting drug-productive clusters.

Although their development started in the post-NP era, 16/21 (76.2%) leads have been initially discovered in the NP era by low-throughput screening (LTS) (8 leads), exploration of known target-binders (ETB), such as hormones/factors and ligands/substrates (5 leads), focused library screening (FLT) of selected structural, target or

Table 1 Statistics of nature-derived FDA-approved drugs in 2008–2012

Year	Total number of FDA-approved drugs	Number (%) of natural products	Number (%) of natural product derivatives	Number (%) of biologics of nonhuman origins	Number (%) of biologics of human origin
2012	37	3 (8.1)	6 (16.2)	0 (0.0)	9 (24.3)
2011	28	2 (7.1)	6 (16.2)	0 (0.0)	5 (13.5)
2010	22	0 (0.0)	9 (24.3)	2 (5.4)	7 (18.9)
2009	25	1 (4.0)	8 (21.6)	2 (5.4)	4 (10.8)
2008	20	0 (0.0)	4 (10.8)	0 (0.0)	3 (8.1)

host species classes (2 leads) and focused bioprospecting of unexplored species of drug-productive species families (1 lead). Five NMEs were discovered in the post-NP era by ETB, FLT, high-throughput screening (HTS) and incorporation of NP component into leads, which suggests that ETB and FLT as well as LTS are highly useful for discovering new NP leads.

The leads of 17/21 (81%) NMEs have one or more of the following deficiencies: unfavorable pharmacokinetic properties (11 leads), insufficient potency (4 leads), lower target selectivity (2 leads) and drug resistance (1 lead). Pharmacokinetic deficiencies typically include low half-life or metabolic stability (8 leads), poor solubility (4 leads), insufficient oral absorption (1 lead) and excessive plasma protein binding (1 lead). The strategies for overcoming some of these deficiencies have been described in the literature reporting the discovery of these drugs. As these deficiencies are quite common in bioactive NPs[7,8], these strategies may be further expanded and more extensively applied in future drug development efforts.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENT

This study was supported by NUS Academic Research Fund (R-148-000-181-112) and NGS Scholarship in Singapore, Chinese Ministry of Science and Technology 863 Hi-Tech Program (2007AA02Z160) and Key Special Project Grant (2009ZX09501-004), and the Fundamental Research Funds for the Central Universities of China.

AUTHOR CONTRIBUTIONS

L.T., Y.Y.J. and Y.Z.C. conceived and managed the project, L.T. and F.Z. developed data search tools, L.T., F.Z., C.Q., C.Z., C.Y.T. and Y.Z.C. searched the data, L.T., F.X., Y.Y.J. and Y.Z.C. analyzed the data, and L.T. and Y.Z.C. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Lin Tao[1–3], Feng Zhu[2], Chu Qin[2,3], Cheng Zhang[2], Feng Xu[4], Chun Yan Tan[1], Yu Yang Jiang[1] & Yu Zong Chen[1–3]**

[1]Shenzhen Kivita Innovative Drug Discovery Institute, Pharmaceutical Sciences, School of Medicine, Tsinghua University, Beijing, P.R. China, and the Ministry-Province Jointly Constructed Base for State Key Lab-Shenzhen Key Laboratory of Chemical Biology, the Graduate School at Shenzhen, Tsinghua University, Shenzhen, P.R. China. [2]Bioinformatics and Drug Design Group, Department of Pharmacy, and Center for Computational Science and Engineering, National University of Singapore, Singapore. [3]NUS Graduate School for Integrative Sciences and Engineering, Singapore. [4]State Key Laboratory of Medicinal Chemistry & Biology, Tianjin International Joint Academy of Biotechnology & Medicine, Tianjin, China.
e-mail: phacyz@nus.edu.sg or jiangyy@sz.tsinghua.edu.cn

1. Lipinski, C. & Hopkins, A. *Nature* **432**, 855–861 (2004).
2. Newman, D.J. & Cragg, G.M. *J. Nat. Prod.* **75**, 311–335 (2012).
3. Zhu, F. *et al. Proc. Natl. Acad. Sci. USA* **108**, 12943–12948 (2011).
4. Overington, J.P., Al-Lazikani, B. & Hopkins, A.L. *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
5. Drag, M. & Salvesen, G.S. *Nat. Rev. Drug Discov.* **9**, 690–701 (2010).
6. Berndt, N., Hamilton, A.D. & Sebti, S.M. *Nat. Rev. Cancer* **11**, 775–791 (2011).
7. Qin, C. *et al. PLoS ONE* **7**, e49969 (2012).
8. Harris, P.A. *et al. J. Med. Chem.* **51**, 4632–4640 (2008).

Development of the clinical next-generation sequencing industry in a shifting policy climate

To the Editor:

Next-generation sequencing (NGS) technologies are increasingly being integrated into clinical practice[1,2]. Proponents expect that the technology will continue to improve clinical care, and early reports suggest some clinical utility in oncology and the diagnosis of rare diseases[3–5]. The industry, however, is in a considerable state of flux as new business models emerge, existing businesses begin to consolidate and the industry reacts to an uncertain regulatory climate[6]. Regulatory standards for test quality and reimbursement remain vague. For example, more than two years ago the US Food and Drug Administration (FDA) asserted its jurisdiction over laboratory-developed tests (LDTs), including some NGS technologies. In November 2013, FDA began regulating NGS machines, granting marketing authorization for the first high-throughput NGS genomics platform (http://blogs.fda.gov/fdavoice/index.php/2013/11/gene-sequencing-devices-are-next-generation/). In the words of FDA Commissioner Margaret Hamburg and National Institutes of Health Director Francis Collins, "This marketing authorization of a non–disease-specific platform will allow any lab to test any sequence for any purpose"[7]. Even so, many questions remain about how and if FDA will regulate NGS-based LDTs, and what clinical validity standards manufacturers may need to meet. In July 2014, the FDA took substantial steps to address these questions by releasing a draft framework for how the agency will regulate LDTs. These guidelines, however, will take upwards of two years to finalize, and five to ten years to implement. Moreover, it remains unclear whether and how they will address issues specific to clinical NGS[8].

Meanwhile, the US Centers for Medicare and Medicaid Services (CMS) have yet to adopt a clear coding and reimbursement structure for NGS in clinical practice, and Blue Cross Blue Shield has recently decided that whole exome and whole genome sequencing should be treated as experimental and not routinely reimbursed[9]. Finally, although the recent US Supreme Court ruling on *Association for Molecular Pathology v. Myriad Genetics*[10] has begun to address uncertainty about intellectual property claims, it has also opened a new set of related lawsuits as well as a broader debate about proprietary genomic databases[11].

Much has been written about these policy gaps, but it is important to understand how industry leaders are responding to the uncertainty and what implications the responses have for future policy development. Here we present the results of in-depth interviews with leaders of 19 companies and laboratories involved in clinical NGS. Our findings indicate that the industry is developing along the NGS pipeline partly