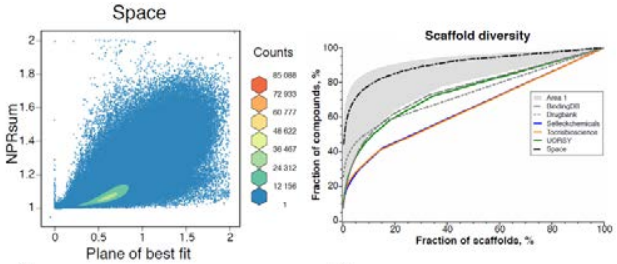
Evolution of commercially available compounds for HTS.

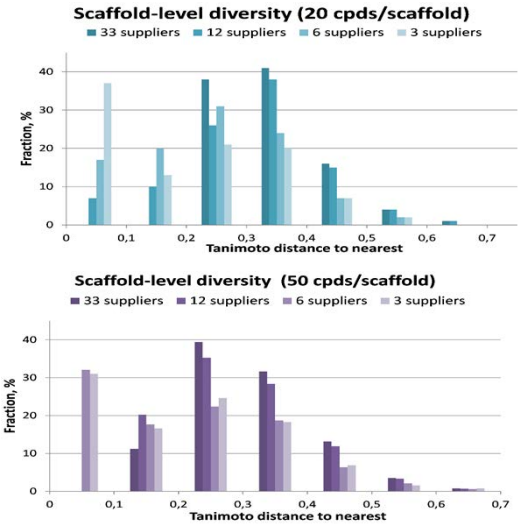
D. Volochnyuk, S. Ryabukhin, Y. Moroz

Introduction and Aim

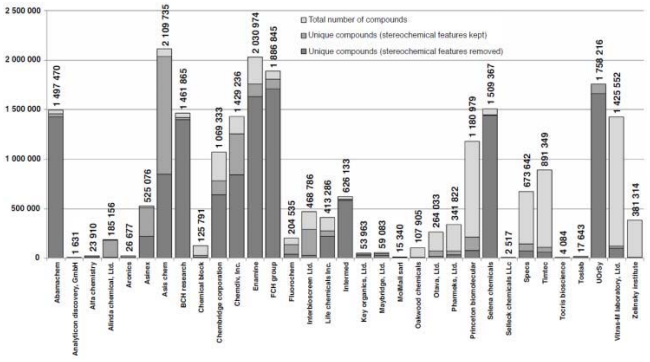
As HTS has matured, the understanding of what features constitute a quality hit and lead has evolved. Over recent years an industry of compound suppliers has grown to provide drug discovery with screening compounds: it's estimated that there are over 16 million compounds available from these sources. It is generally regarded that low lipophilic, and higher fsp3 properties are preferred. This paper reviews the chemical space covered by suppliers' compound libraries (SCL) in terms of compound chemophysical properties, novelty, diversity and quality. It examines the feasibility of compiling high-quality vendor-based libraries avoiding complicated, expensive compound management activity and compares the resulting libraries to the ChEMBL dataset. From our analysis it would appear that over the last 10 years the market has evolved to meet these demands, with new compounds from many suppliers meeting modern physiochemical properties. At the moment it is not possible to purchase an ideal one-million compound set (50K scaffolds, minimum of 20 compounds per scaffold). However, it would appear that an ideal 500K set can be purchased. If sample logistics is an issue then we have shown that it is possible to purchase the 500k set from only six suppliers, with a 350K set available from just three suppliers. Many large companies have been through similar exercises and have built their screening decks accordingly. If you are considering building a screening deck Ab initio then it is possible to achieve this from purchasable space. On the basis of this review we are confident that as new challenges in sample supply emerge then the market place will respond positively.



(a) Density plot of Plane of Best Fit (PBF) score versus the sum of normalized principal moments of inertia (NPR). (b) Cumulative Scaffold Frequency Plots of the scaffold with 'vendor areas' and outliers compared with Binding DB and DrugBank.



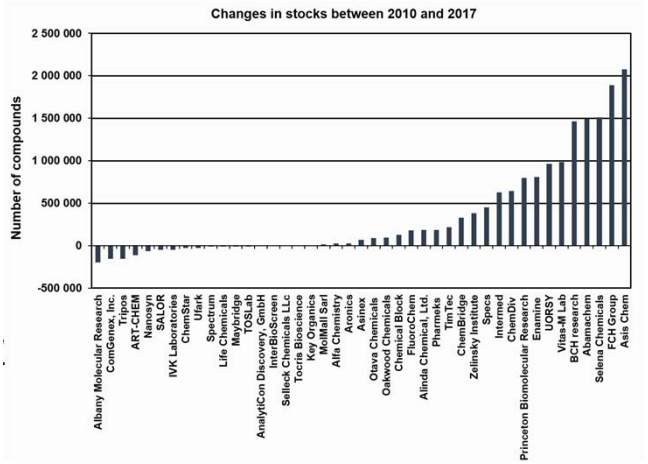
Comparison of the scaffold diversity of the libraries collected from 33, 12, six, and three suppliers. (a) For 20 compounds per scaffold set; (b) For 50 compounds per scaffold set.



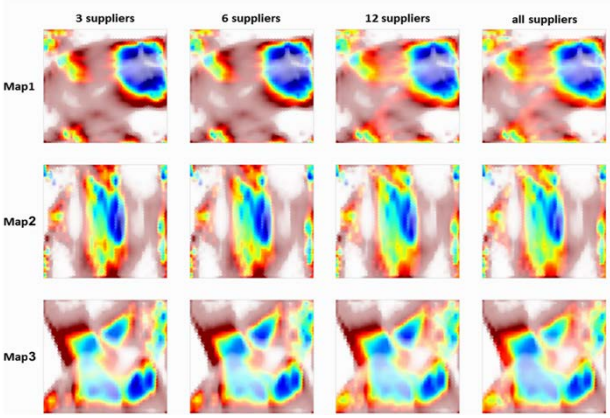
The chemical space of purchasable screening compounds represented by vendors.

Mean values of selected molecular properties of the purchasable chemical space in 2010, 2017 and ELF library.

Parameter (X)	2010	2017	Δ<X> (2010–2017)	<X>Δ (2010–2017)	ELF
MW	388.82	362.49	−26.33	339.59	425
logP	3.64	2.96	−0.67	2.38	3.1
Fsp³	–	0.40	–	–	0.4
tPSA	94.23	71.84	−22.39	52.38	91
Heavy atoms	–	25.11	–	–	–
HBA	6.18	4.61	−1.57	3.25	–
HBD	0.96	1.16	0.20	1.33	–
ROTB	5.28	4.82	−0.47	4.41	–
Rings	–	3.02	–	–	–
Aromatic rings	–	2.03	–	–	–



Changes in suppliers' compound libraries (SCL) size from 2010 to 2017.



Generative Topographic Mapping (GTM) maps of four compound sets corresponding to three, six, 12, and 33 suppliers on the ChEMBL compounds background. See main text for key to colors.

Contacts

Dmitriy M. Volochnyuk, Prof. Dr. Sci; Sergey V. Ryabukhin, Prof. Dr. Sci.
d,volochnyuk@gmail.com, s.v.ryabukhin@gmail.com.
Enamine Ltd, www.enamine.net 78 Chervonotkatska St, 02660 Kyiv, Ukraine

References

1. Volochnyuk D.M., Ryabukhin S.V., et al. *Drug Discovery Today* **2019**, 2, 390-402.