Identifying Candidate Ataxin-2 Inhibitors in High-Throughput Screening Data using Molecular Descriptors

David E. Jones [1], Jingran Wen [1], Smita Sahay [2], Daniel R. Scoles [3], Thomas Dexheimer [4], Hongmao Sun [4], David Maloney [4], Anton Simeonoy [4], Ajit Jadhay [4], Stefan M. Pulst [3], and Julio C. Facelli [1]

[1] Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; [2] Department of Bioengineering, University of California, San Diego, La Jolla, CA; [3] Department of Neurology, University of Utah, Salt Lake City, UT; and [4] National Center for Advancing Translational Sciences (NCATS), Rockville, MD

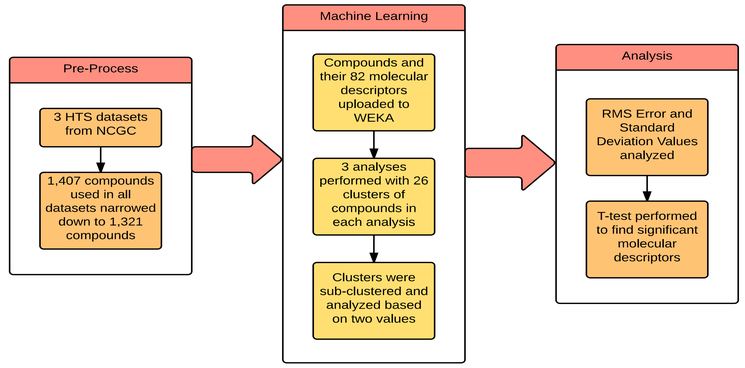
Background

- Spinocerebellar Ataxia Type II (SCA2) is an autosomal dominant neurodegenerative disease characterized by cerebellar dysfunction.
- SCA2 is caused by abnormal expansion of CAG repeats in the *ATXN2* gene which gives rise to the abnormal polyglutamine region in the *ATXN2* protein.
- Potential therapeutics focus on decreasing the level of mutant protein by suppressing the expression of the corresponding gene.
- Drug repurposing is a strategy where new clinical trials of drugs known to treat one disease positively show that they can be reused in an effective way for treating other diseases.
- Goal is to analyze molecular descriptors of chemical entities found from the NCGC Pharmaceutical collection to find promising drugs that can be repurposed to treat SCA2.

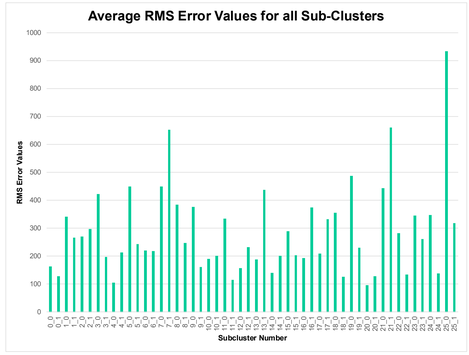
Purpose

In this presentation we demonstrate how high-throughput screening (HTS) and molecular descriptor data was analysed using data mining and machine learning techniques to identify promising drugs for repurposing and treating SCA2.

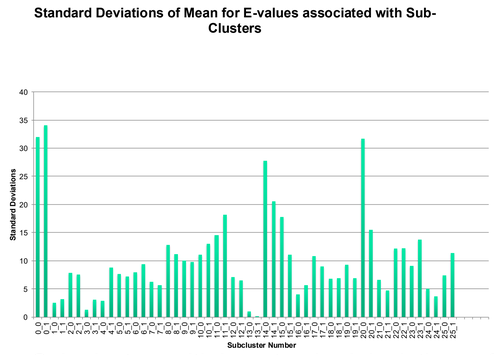
Methods



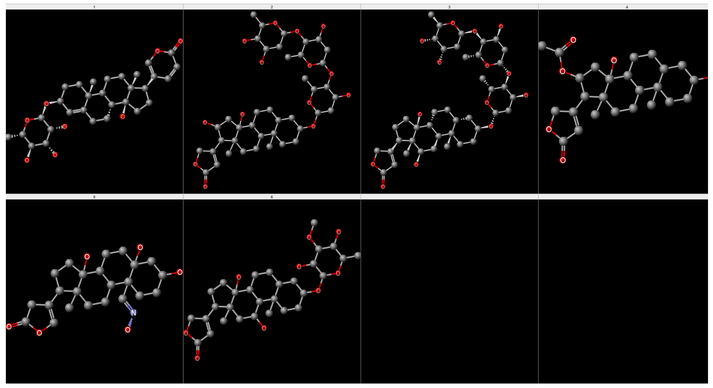
Results



Sub-clusters with low RMS error values contained compounds that shared structural properties.



Sub-clusters with low standard deviations of the mean for the E-value contained compounds that shared *ATXN2* inhibition qualities.



Visual representation of 6 compounds from sub-cluster 4.0 that are good inhibitors of *ATXN2.*

Molecular Descriptor	P-Value
Aliphatic Ring Count	0.0202
Aromatic Atom Count	0.0167
Aromatic Bond Count	0.0153
Aromatic Ring Count	0.0133
Asymmetric Atom Count	0.0101
Carboaliphatic Ring Count	0.0016
Carboaromatic Ring Count	0.0195
Carbo Ring Count	0.0301
Chiral Center Count	0.0134
Aliphatic Ring Ratio	0.0202
Aromatic Atom Ratio	0.0182
Aromatic Bond Ratio	0.0171
Aromatic Ring Ratio	0.0197
Asymmetric Atom Ratio	0.0057
Carboaliphatic Ring Ratio	0.0052
Carboaromatic Ring Ratio	0.0333
Chiral Center Ratio	0.0065

P-values for the statistically significant molecular descriptors.

Discussion

- Two sub-clusters had been determined to be of interest. One sub-cluster contained compounds which have the greatest potential to specifically inhibit the downstream products of *ATXN2*. The other sub-cluster contained compounds which have low potential to specifically inhibit the downstream products of *ATXN2*.
- A t-test between these two sub-clusters revealed that the molecular descriptors of compounds that could potentially inhibit the downstream products of *ATXN2* from the provided NCGC HTS data were aliphatic ring count, asymmetric atom count, carboaliphatic ring count, carbo ring count, and chiral center count.

Conclusion

Machine learning and data mining can be used effectively to analyze data from HTS for the repurposing of drugs. The method used here is able to clearly identify candidate inhibitors for treating SCA2.

Acknowledgements

DEJ and SS have been supported by NLM training grant 5T15LM007124, JW was partially supported by the Fay's Informatics Research Fellowship. The authors acknowledge NIH grants RC4NS073009, R21NS081182, and R01NS033123 (DRS and SMP), NCATS U54MH084681 (TD, HS, DM, AS, AJ) and NLM 5T15LM007124 and NCATS 1ULTR001067 (JCF).

Contact Information

David E. Jones
davide.jones@utah.edu

The University of Utah
Biomedical Informatics