Working title:

Automated SAR-based triage: lessons from applying the NIH R-group tool in pharma

<author list: may add a couple GSK/NIH authors as needed>

Abstract:

<TBD>

Introduction:

<TBD>

* Related work:
  + U. Bonn/Bajorath et al.: activity cliffs/landscapes, SAR rules (M. Wawer), SAR index, polypharmacology, …
  + Novartis/E. Lounkine, A.M. Wassermann et al.: (navigating chem.. space, side effect pred.)
  + Novartis/A. Schuffenhauer (Scaffold networks, Latent hits; include SNG by S.J. Swamidass)
  + GSK/Frameworks, Reduced Graphs and Data-Driven Clustering (G. Harper et al., JCICS 2001)
  + Other references to standard fingerprint-based or other clustering techniques, or a review
  + Discuss notion of privileged scaffolds and their relevance / influence on scaffold based triage
  + Anything previous out of NIH … (please help me out here?)
  + Some refs to applications papers mentioning how HTS datasets were triaged using various methods
    - Peng et al “A Crowd-Based Process and Tool for HTS Hit Triage”
    - Cox et al “Abbott Physicochemical Tiering (APT)—A unified approach to HTS triage”
    - Many more to pick from

Methods:

1. NIH R-group tool – described in preceding paper
   1. Since we won’t have an explicit paper on the tool before this comes out (?), we should probably have a brief section describing the underlying method
2. Interface from NIH tool to Spotfire
3. Developing a Spotfire UI to enable SAR-based triage

Applications:

1. Scaffold-based triage and progression
   1. Scaffold and R-group statistics/summaries help make the decision to progress a lead with SAR context, highlighting the best lead series regardless of potency or # exemplars
   2. Lead-likeness. Visualize whether activity is linked to undesirable properties such as high PFI or # aromatic rings.
   3. …
2. Automated R-group table generation can be a mini-application in itself
   1. Replacing manual effort that is expended before every chemistry meeting
3. Merging hit ID datasets by chemotype
   1. Finding shared/unique chemotypes
   2. Designing hybrid analogs by combining SAR from multiple datasets
4. Scaffold walking to navigate between hit compounds
   1. Deconstruct a compound into all scaffolds it contains
   2. Link from a compound to the SAR tables/plots for each of its scaffolds
   3. 🡪 application: Identify more potent/ligand-efficient compounds by “scaffold walking” from a known hit
   4. 🡪 capability: automated/trivial substructure search within dataset
5. …

Discussion:

<TBD>

* + Enumerate challenges and future directions from the applications perspective, taking care to keep them separate from this same section in the methods paper, which would focus on pure methodological improvements.

References:

<TBD>