Team name	VirtualFlow@Covid19
Team member(s) (firstname lastname; ...)	Christoph Gorgulla Krishna M. Padmanabha Das Kendra E. Leigh Marco Cespugli Guilhem Tesseyre Shreya Pandita Alec Shnapir Vedat Durmaz Konstantin Fackeldey Christian Gruber Ryan Yust Dave Payne Gerhard Wagner Haribabu Arthanari Noam Lewis Colin Hutcheson Erez Yaffe
Affiliation	Harvard/DFCI, Google, Innophore, TU Berlin
Contact email	cgorgulla@g.harvard.edu
Contact phone number (optional)	+18574150289
Protein targets (for example: 3CLPro/Nsp5, BoAT1, Fc Receptor, Furin, IL6R, M protein, Nspx, OrfXx, N, E, etc...) | 3 required	Mpro PLpro nsp12 Spike nucleoprotein (N) TMPRSS2

Section 1: methods & metrics

Describe what methods you have used, how they are independent from one another, what your workflow was, how you performed the cross-correlation between your methods. If applicable, please report estimated performance metrics of your methods, such as accuracy, sensitivity, false-discovery rate, etc., and how those metrics were obtained (e.g. cross-validation). Please provide key references if available.

Methods:

We used VirtualFlow for all targets [https://www.nature.com/articles/s41586-020-2117-z].
We carried our rigid docking for all targets.
This was the only method we applied.

We filtered the hit lists with DataWarrior [http://www.openmolecules.org/datawarrior/].

Section 2: targets

Describe for each protein target: why you chose it, from which source you obtained it (e.g., insidecorona.net / covid.molssi.org / rcsb.org) and why this is the best quality structure, if any pre-processing (e.g., energy minimization, residue correction, alternative folding, …) was performed.

Target 1: Mpro

- *Reason: Vital protein; structure available;*
- *Structure used: 6lu7*
- *Quality: Structure was peer reviewed and published in Nature. At the time we did the screening (early) it was one of the few structures available.*

Target 2: PLpro

- *Reason: Vital protein; structure available;*
- *Structure used: 6w9c*
- *Quality: Good resolution and fully resolved. At the time we did the screening (early) it was one of the few structures available.*

Target 3: nsp12

- *Reason: Vital protein; structure available;*
- *Structure used: 7BV1*
- *Quality: Good resolution and fully resolved at the relevant parts. Quality: Peer-reviewed, published in Science*

Target 4: Spike

- *Reason: Vital protein; structure available;*
- *Structure used: 6W41*
- *Quality: Peer-reviewed, published in Science*

Target 5: nucleoprotein (N) - CTD

- *Reason: Vital protein; structure available;*
- *Structure used: 6WJI*
- *Quality: Good resolution, fully resolved CTD*

Target 6: TMPRSS2

- *Reason: Important protein for the virus*
- *Structure used: Swissmodel homology model*
- *Quality: High quality according to Swissmodel*

Section 3: libraries

Describe which libraries you have used, how they were combined, if any compounds were removed / added, why additions are relevant, any unique features of your library, etc. Please provide the sources you obtained the libraries from (if publicly available). Describe the procedure of data preparation (removal of duplicates, standardization, etc). Indicate if different libraries were used for different targets, and why. If possible, provide a download link to your version of the library.

Library 1:

REAL library from Enamine, containing 1.4 billion molecules. We have screened around 1 billion for each target.

https://virtual-flow.org/real-library
https://www.nature.com/articles/s41586-020-2117-z

Library 2:

ZINC library: Around 10 million in stock compounds. Contains most of the SWEETLEAD compounds, so we have not screened the SWEETLEAD library separately.

Section 4: results

Briefly describe you key findings, any interesting trends in your data, a description of your top 5 compounds for each target. If possible, provide a link to a code and/or data repository. Please do not submit randomly selected compounds!

Results:

The top five compounds, which are at the top of our lists, have nice docking poses, and are mostly beautiful in terms of MedChem. They mostly have no predicted toxicity or reactive groups.