Shifted Transversal Design smart-pooling for high coverage interactome mapping

Xiaofeng Xin, Jean-François Rual, Tomoko Hirozane-Kishikawa, David E Hill, Marc Vidal, Charles Boone, Nicolas Thierry-Mieg

Running title: STD pooling for high coverage interactome mapping

Supplementary Material Content:

Supplementary Note

Pilot experiment: comparison with a subset of CCSB-HI1

Prior to the *C. elegans* ORFeome-wide experiments, we studied the feasibility of an STD-based smart-pooling strategy in the context of Y2H interactome mapping using the CCSB-HI1 dataset (Rual et al. 2005) as a benchmark. We picked 100 baits (**Supplementary Table 5**) and 940 preys from the Human ORFeome (10 Human ORFeome AD plates: YAH11003, 11006, 11008, 11013, 11014, 11017, 11020, 11024, 11025 and 11086), covering a varied subset of the previously screened CCSB-HI1 space. The subspace was biased to include highly-connected baits as well as some known auto-activators, in order to assess STD pooling in the most challenging conditions. Indeed, hubs are the known weak spot of smart-pooling, as decoding may become ambiguous and require sequencing to identify some of the positives, and auto-activators are notoriously problematic in Y2H. After performing extensive simulations using the interpool software as described (Thierry-Mieg and Bailly 2008), we selected the design STD(940;13;13) (see Thierry-Mieg 2006 for details). In this design, the 940 preys are arrayed into 169 pools comprising 72 or 73 preys each, and each prey is present in 13 pools. Three pools uniquely identify a prey, leaving an extra redundancy of 10 pools for correcting noise and dealing with multiple positives. The relatively small pool size (for a 96-format assay) was chosen for this feasibility study in order to avoid potential issues due to excessive dilution, although the resulting design is not competitive in terms of workload. The pools were assembled using robotics similarly to the worm STD pools (see **Methods** online) but without generating intermediate micro-pools. The 169 pools were then screened as two 96-well plates against each of the 100 baits, following the standard Screen-Seq Y2H protocol (Rual et al. 2005) except that we skipped the counter-selection of the cycloheximide marker: this step is designed to detect and eliminate *de novo* auto-activators, but since these are not robustly reproducible we expected that they would be intrinsically avoided by STD pooling. Finally, instead of sequencing positive colonies, the observed patterns of positive spots were reported using a custom web-based tool. Spots were scored using two discrete levels as either positive or negative, and results were decoded with an early version of interpool (Thierry-Mieg and Bailly 2008). Briefly, this algorithm identifies the putative positive preys

most likely to explain the observed pattern, as well as the putative false positive and false negative spots, as illustrated in **Supplementary Figure 2**. In this pilot experiment we used the standard Hamming distance (Ω={neg,pos}, $\delta_{NEG}=\delta_{POS}=1$). Over the course of this work, it became clear that additional discrete levels would be useful. This led to the generalized interpool formulation and algorithm which was described in Thierry-Mieg and Bailly 2008 and used in the worm STD smart-pooling experiments (see **Methods** online). All "novel" interactions (absent from CCSB-HI1) were retested using stringent multi-phenotype testing as previously described (Rual et al. 2005).

Overall 65 putative interactions were identified (**Supplementary Table 6**), among which 60 passed pairwise retest successfully, 3 failed, and 2 could not be confirmed because of auto-activation in the retest, despite a clear signal in the STD pooling screen. These latter 2 hits were not included in the list of confirmed positives, but they illustrate that auto-activation is not an all-or-nothing process. Depending on timing and small changes in experimental conditions, hits can emerge more or less over background growth, and STD pooling provides a way of discerning between signal and noise. Notwithstanding, this demonstrated the high specificity of the STD pooling strategy, with a 92% Positive Predictive Value (PPV; the percentage of hits that turn out to be true positives). The 60 confirmed hits include 31 novel interactions and 29 found in CCSB-HI1, representing 73% of the 40 CCSB-HI1 positives in the explored subspace. It is difficult to estimate the sensitivity of STD pooling from this experiment, because the subspace was chosen to include many CCSB-HI1 interactions, but the 31 novel interactions suggest a sensitivity between 1.5 and 2.8 times higher than the standard Screen-Seq protocol (see **Supplementary Methods** online), while the specificity is excellent.

Supplementary Methods

Construction of dilution plates for pool size testing

1. Eight Worm ORFeome AD plates (plate numbers appear in **Supplementary Table 2c**) were inoculated in 96-format deep well plates containing SD-Trp media, incubated at 30°C for 2 days.
2. For each ORFeome 96-well plate used in this dilution test, every 4 wells were mixed together to make a pool of 4 preys according to **Supplementary Table 2a** (P24 contains 2 preys because G12 and H12 are empty in worm ORFeome plates). For example, A1, B1, C1 and D1 were pooled to make pool P1, and E1, F1, G1 and H1 were pooled to make pool P2. In this way, 24 4-prey pools were generated from one 96-well plate.
3. The 4-prey pools were pooled together to make a series of pools of increasing sizes ranging from 4 to 94, named W1 to W23, as specified in **Supplementary Table 2b**. The pools are nested: W(n) contains all W(n-1) preys and another 4 additional preys from P(n) (where 1< n <= 22). W1 is P1, W23 contains W22 plus P23 and P24, W24 contains no prey as a negative control. The 24 W-pools were arrayed into one W-row in a 384-well plate. Thus, from one 96-well ORFeome plate, we generated one row (in a 384-well plate) of nested pools whose sizes increase gradually by increments of 4 (6 for W23).
4. Each W-row from step 3 was placed in duplicate rows on the final 384-well dilution plate according to **Supplementary Table 2c**.
5. The resulting 384-well dilution plate was transferred to an SD-Trp omnitray agar plate using a "BM3-SC+Carousel" robot (S & P Robotics) after the yeast cells in each well were resuspended and mixed well.
6. The 1536-format dilution agar plate was generated from the 384-format dilution agar plate generated above by using a "BM3-SC+Carousel" robot (S & P Robotics), pinning each spot in quadruplicate (as illustrated at the bottom-left corner of the 1536-format plate in **Supplementary Fig. 1**).

Pilot experiment sensitivity

There are 40 CCSB-HI1 interactions in the 100x940 subspace explored in this pilot experiment, while we found 60 positives. This suggests a 1.5-fold increased sensitivity of STD pooling compared to the CCSB-HI1 protocol. However, the explored subspace was strongly biased to include a large number of CCSB-HI1 interactions. This 1.5-fold increase is therefore an underestimate. A better estimate can be derived by excluding the interactions that were found by both methods. CCSB-HI1 then contains 11 hits while this pilot experiment identified 31 novel interactions, representing a 2.8-fold increased sensitivity. Another approach consists in considering the remaining 100x7160 subspace of CCSB-HI1 corresponding to the chosen baits, where 94 interactions were found. Scaling down to the pilot experiment subspace size, we would therefore expect ~12.3 CCSB-HI1 interactions. Hence our 31 novel positives represent a 2.5-fold higher sensitivity, in general agreement with the previous estimate.

Supplementary Figure Legends

Supplementary Figure 1 | Example of dilution test plates for identifying the largest usable pool size in each format. See **Supplementary Methods** and **Supplementary Table 2** for the description of the dilution plate construction. **(a)** 384-format test plate, with the SH3 domain of *C. elegans* Y106G6H.14 (10-86 a.a.) as bait. **(b)** 1536-format test plate, with the SH3 domain of *C. elegans* F09E10.8a (520-592 a.a.) as bait. Numbers above the plates show the size of the pools in each column. The red arrows and numbers indicate the pool size chosen for each format. Numbers on the left are the worm ORFeome plate IDs that were used to generate the corresponding rows. For 384- and 1536-format plates, each spot is in duplicate and octuplicate, respectively, as illustrated at the bottom left corner (red boxes). Each blue box highlights a series of pools (of increasing size from left to right) containing a given interactor. The effect of dilution is more obvious on the 1536-format plate, particularly for the rows from ORFeome plate 11021. Yellow boxes show mixed large and small spots in a series of pools, which may result from stochastic events in Y2H (e.g. very weak PPIs which can only be detected when the number of plasmids carried by yeast is above a certain threshold).

Supplementary Figure 2 | Example pilot experiment screen result (see **Supplementary Note**) for bait 11052@C05. The 169 pools are arrayed on two 96-well plates. Spots scored as positive are circled (no color at this stage). Decoding with interpool identifies two putative positive preys: 11006@G08 (green circles), which has no false negative pools and was previously identified in CCSB-HI1; and 11025@H01 (blue circles), a novel interactor for this bait. Three spots containing 11025@H01 were not scored as positive and are proposed as false negatives by interpool (red boxes), and two spots are identified as false positives (red circles). Despite the fact that the signal is quite weak for all spots containing 11025@H01, it is unmistakably identified by its 10 positive spots, illustrating the usefulness of highly redundant pools.

Supplementary Figure 3 | **(a)** Sensitivity and **(b)** PPV of STD-1536 and STD-SL restricted to the 3 hub baits screened in STD-SL. Error bars indicate standard error.

Supplementary Figure 4 | Repeatability of **(a)** Core and **(b)** FP hits from each ORFeome-wide Y2H dataset. Each screen was performed twice; the percentage (and total number) of each method's hits that were identified in both replicates is displayed.

References

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N. et al. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437:** 1173-1178.

Thierry-Mieg, N. and Bailly, G. 2008. Interpool: interpreting smart-pooling results. *Bioinformatics* **24:** 696-703.

List of Supplementary Tables

Supplementary Table 1: List of *C. elegans* baits used in the ORFeome-wide experiments.

Supplementary Table 2: Table (with 3 sub-tables) to illustrate the construction of dilution plates for pool size testing.

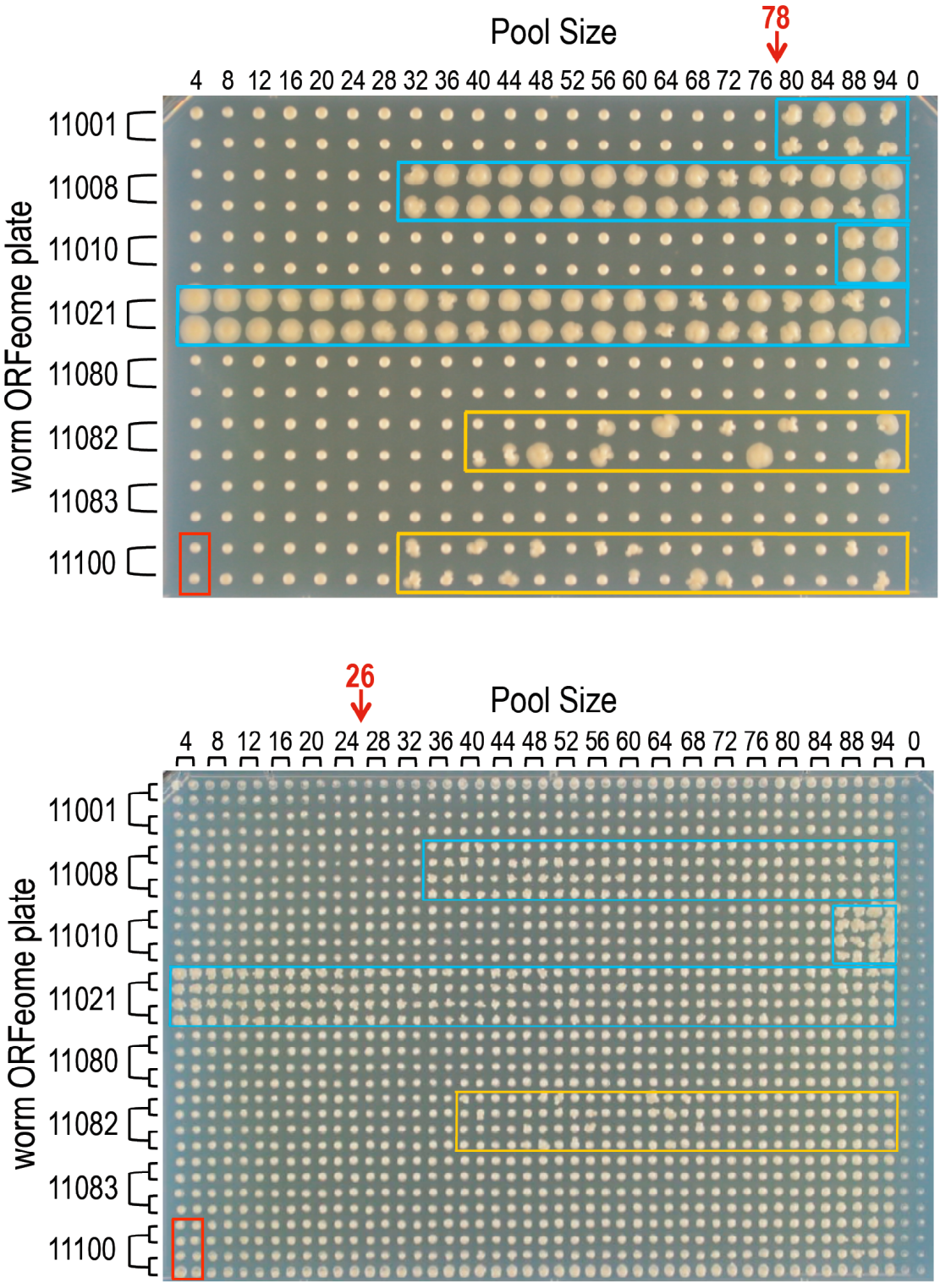
Supplementary Table 3: NonCore interactions detected by each method in the *C. elegans* ORFeome-wide experiments (Worm ORFeome IDs).

Supplementary Table 4: Core interactions detected by each method in the *C. elegans* ORFeome-wide experiments (Worm ORFeome IDs).

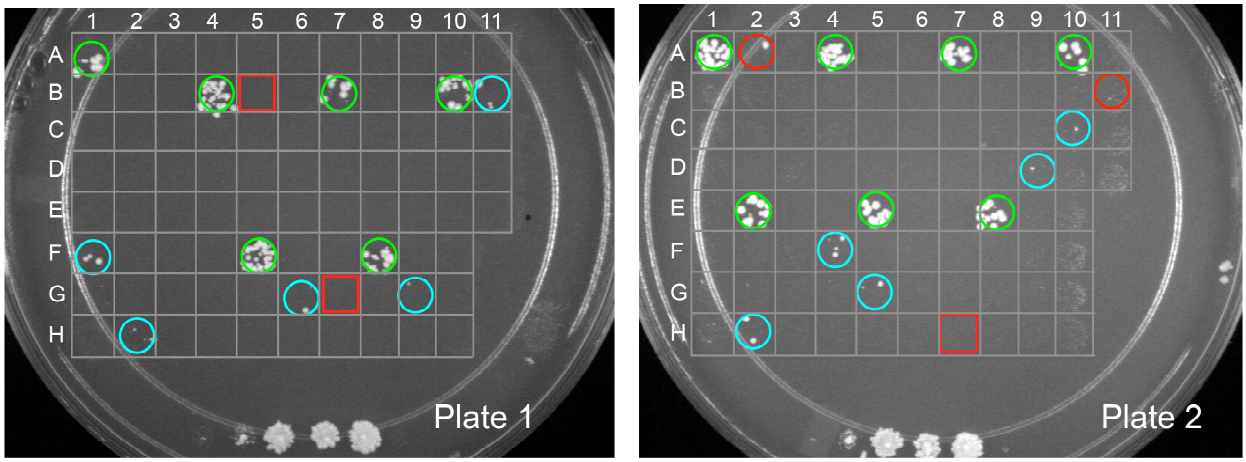
Supplementary Table 5: List of Human ORFeome baits used in the pilot experiment.

Supplementary Table 6: Interactions detected in the pilot experiment (Human ORFeome IDs).

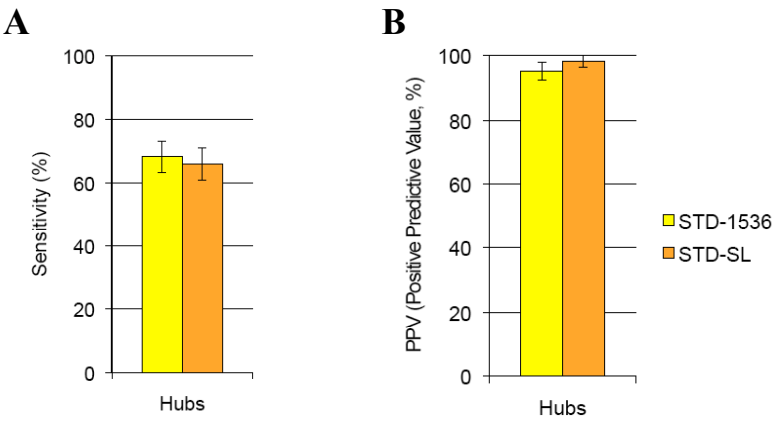
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4