The goal of this project is to reveal differences in protein associations present in healthy versus diseased human tissues, providing targets for diagnostics and therapeutics; protein-protein interfaces are increasingly recognized as a large, important body of druggable targets. Antibodies are the most commonly employed protein detection and capture reagents used in research and diagnostic laboratories. Monoclonal antibodies (**mAbs**) are also increasingly used as therapeutics. Antibodies can facilitate the study of the associations between different proteins in the cell using a technique known as immunoprecipitation (**IP**). However, it is widely recognized that the lack of high quality antibodies against human proteins is negatively impacting biomedical science. The market for research antibodies alone is ~$2.5 billion annually, but estimates of capital losses due to poor quality antibodies amount to $800 million per year. Improving the *quality* and the *rate* of antibody production is critical. To improve quality we will establish rigorous validation of mAbs for their ability to IP endogenous protein complexes; to accelerate production we have preliminarily demonstrated that endogenous protein complexes purified from cultured human cells can serve as efficient immunogens for generating multiple mAbs at once.

CDI Laboratories Inc. has developed a proprietary pipeline employing the largest-content full-length human protein (**HuProt**) microarray in existence, allowing them to generate mouse mAbs that exhibit *monospecificity* for target human proteins. CDI already has >500 array-validated mAbs against human transcription factors (**TF**) – a well-recognized class of disease-linked proteins. *Dr. John LaCava* has developed *a screen that enables and optimizes the capture of endogenous protein complexes by IP* in conjunction with mass spectrometry (**MS**)-based interactome analyses. In Phase I we will (1) validate CDI's existing mAb catalog for IP of endogenous disease-linked TF protein complexes and, in an independent aim, (2) increase the production of new high performance mAbs using endogenously assembled epitope-tagged protein complexes. In Phase II, IP of native endogenous complexes using mAbs will be combined with the generation of novel mAbs from the resulting complexes, foregoing tagging and further accelerating production to rapidly increase the coverage of mAbs against the human proteome. This will simultaneously allow *mining the interactome to identify pathogenic alterations that may serve as diagnostic and therapeutic targets*. The outcome of Phase I will be of a collection of application-validated mAbs for research and diagnostics.

Specific Aims

Aim 1: Characterize CDI's anti-human TF mAbs for their ability to IP endogenous cancer-related TF protein complexes. *Hypothesis: mAbs from CDI's extant anti-human TF collection can be used to IP TF protein complexes as expressed and assembled natively in tissue culture cells and patient-derived samples.*

a) mAbs that target TFs expressed in HEK and/or HeLa cells will be screened for their ability to IP protein complexes from these cells. mAbs against TF targets that are cancer-related will be prioritized.
b) mAbs that IP discrete complexes during screening will be characterized by quantitative-MS in up to three associated working conditions. mAbs yielding the target and at least one bona fide interactor within the top five most abundant proteins will be categorized as 'endogenous-complex-IP-competent;' the IP conditions will be ranked according to the number and abundance of bona fide interactors.
c) Newly characterized endogenous-complex-IP-competent mAbs will be tested for their ability to IP TF protein complexes from patient derived cancers obtained through our clinical collaborators.

Milestones: (1) quantitatively characterize the endogenous-complex-IP-competency of at least ten of CDI's anti-human TF mAbs; **(2)** validate up to five endogenous-complex-IP-competent mAbs on patient cancers.

Aim 2: Generate novel mAbs from purified endogenous TF protein complexes. *Hypothesis: endogenously assembled protein complexes captured by IP from cultured human cells can be used as effective immunogens for rapidly producing collections of novel mAbs against multiple complex members.*

a) Cancer-related TF targets that are expressed in HEK and/or HeLa cells will be epitope-tagged with 3xFLAG and subjected to IP optimization screening. CRISPR-based methods will be used to create cell lines expressing TFs tagged at endogenous genomic loci, expressed from their native promoters.
b) TF protein complexes discovered and optimized during IP screening will be scaled-up and prepared for injection into mice to produce collections of mAbs against the constituents of the purified fractions.
c) Targets of novel mAbs will be deconvoluted using CDI's HuProt microarray, and a selected set of monospecific mAbs, with target specificities ≥ 3 standard deviations above any nonspecific target, will be subjected to endogenous-complex-IP-competency testing and characterization (as in Aim 1a, b).

Milestones: (1) workup at least five TF protein complexes with constituent proteins for which mAbs do not currently exist; **(2)** generate up to fifty novel mAbs from TF protein complex immunizations.

Research Strategy

Significance – We will (I) establish procedures to validate monoclonal antibodies for use in the purification of endogenous human protein complexes from cell lines and patient-derived cancers, including metrics for describing the quality of their performance and (II) increase the efficiency and quality of production of new mononclonal antibodies against endogenous human proteins. These two main Aims will together contribute a solution to the ~$800 million per year in capital losses associated with poorly performing research antibodies (2) and improve our understanding of the protein-interaction-level effectors of cancer. Our starting point will be analysis of cancer-linked human transcription factor (**TF**) protein complexes.

*Cellular functionality is chiefly mediated through protein-protein interaction (**PPI**) networks*, the study of which enables understanding of the molecular biology of the cell, including pathogenic network alterations that lead to cancer and disease (3,4). Genome studies alone have not comprehensively revealed the features distinguishing health from disease nor delivered on the promise of transforming medical practice (5). Because protein complexes are the effector molecules of cellular processes, new research efforts are frequently being directed instead towards e.g. defining cancer-related changes in PPI networks (**interactomics**). Identifying disease-related alterations in PPIs has revealed diagnostic and prognostic markers (6-8) as well as therapeutic targets (8-12). However, most efforts to describe candidate protein complexes, as disease markers or targets, come in the form of predictions based on informatics and data mining of noisy datasets (13-15). Direct validation of these predictions in the form of physically purified complexes is infrequent. However, it is essential that such complexes be purified in order to assess the mechanistic aspects of interaction network alteration and disease manifestation, making protein interactions 'druggable targets' that can be exploited through a number of independent strategies including e.g. small molecules (16-21), antibodies (8,22-24), and RNA interference (25).

TFs and their coregulators will be the focus of this study: the PPIs of these master-regulators of gene expression are significantly under-characterized and the topic of intensive ongoing research; *although TFs number in the thousands, just three such TF genes account for more citations in the literature than all the remaining TFs combined* (26,27); yet it is already well understood that they have significant influence on human health and disease, and are considered effective therapeutic targets (28-30). Consider the following example: the tumor suppressor gene *TP53* is the most frequently mutated gene in cancer (31), coding for the most intensively studied and well-known transcription factor, p53 (26,32-34). Decades of effort have been invested into dissecting the molecular mechanisms of mutant p53 oncoproteins and have revealed diverse modes of action across different interaction networks, including both direct and indirect control over gene expression; despite this, the full breadth of cancer-related mutant p53 PPIs nevertheless remains an open question (33,34), underscoring the difficulty in fully characterizing protein interaction networks. As the vast majority of disease- and cancer-linked TFs have not been studied nearly as well as p53, it stands to reason (and is quantitatively understood (4)) that a significant number of new

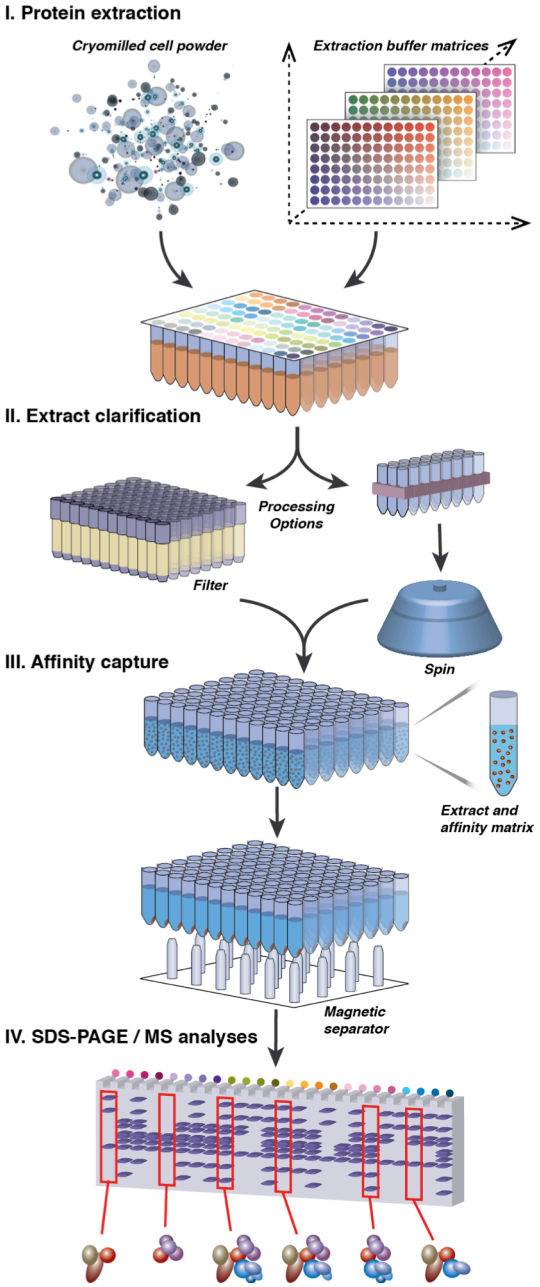

Fig. 1 – *IP optimization procedure* (1)**. (I)** Cells of interest are cryo-milled, and precise aliquots of the resulting frozen powder are deposited into a multiwell plate using a dispensing manifold. A diverse set of extraction solvents is rapidly added in parallel, and the powder is resuspended completely by mechanical agitation and brief sonication. **(II)** Rapid removal of insoluble material is achieved either by centrifugation or using a patented deep-bed filtration device. **(III)** Magnetic bead–based IP of the target protein is performed on the clarified extract using CDI's monoclonal antibodies (for native protein capture; **Aim 1a**) or a commercial antibody against the FLAG epitope (for tagged protein capture; **Aim 2a**). The magnetic medium is collected, washed and eluted with the aid of a multiwell magnetic separator. **(IV)** Elution of the protein staining and analysis by mass spectrometry (**MS**). The resulting copurification profiles are cross-compared to determine relative differences in interactomes. Selected conditions will be chosen for rigorous MS-based quality assessment of the antibody (**Aim 1b**). Conditions presenting a high yield of endogenous tagged protein complexes will be scale-up and entered into CDI's pipeline to produce novel antibodies against the collection of co-purifying proteins (**Aim 2b, 2c**; **see Fig. 2 and *Approach***).

and important mechanistic insights will be gained through their further study. Importantly, some examples of anti-tumorigenic treatments resulting from the interruption of p53 PPIs already exist (35) – illustrating that, if we thoroughly mine disease-linked PPI networks, diagnostic and therapeutic strategies will emerge.

Immunoprecipitation (**IP**) has been the predominant workhorse of PPI studies, with the aim of identifying all of the macromolecular protein complexes formed within the cell (**the interactome**). However, the existing PPI datasets are far from complete (4) owing in part to a lack of IP-competent antibodies. The estimated capital losses due to poor quality antibodies amounts to ~1/3 of the ~$2.5 billion annual market value (2,37). The low quality of many extant antibodies against human proteins is severely limiting biomedical research both in efficiency and in scope, culminating in a community-wide consensus that this problem must be addressed (2,37-41); improving the *rate* and the *quality* of antibody production is of paramount importance. A second major impediment to progress is the need to empirically determine purification conditions permitting the isolation of physiological complexes (1), requiring periods of long trial-and-error experimentation. This proposal concerns initiating a prototyping Phase to create a cyclical pipeline addressing the above-described barriers to progress. We will fuse two independent pipelines, one for IP optimization / interactomic analyses (**Fig. 1**) and one for monoclonal antibody production (**Fig. 2**), to work together in tandem. As a test bed we will initiate our work using two independent methods of IP, (I) using a selection of monoclonal antibodies against cancer-linked human TFs drawn from CDI's extant collection (**Aim 1**), and (II) by epitope tagging selected TFs in model cell lines (**Aim 2**).

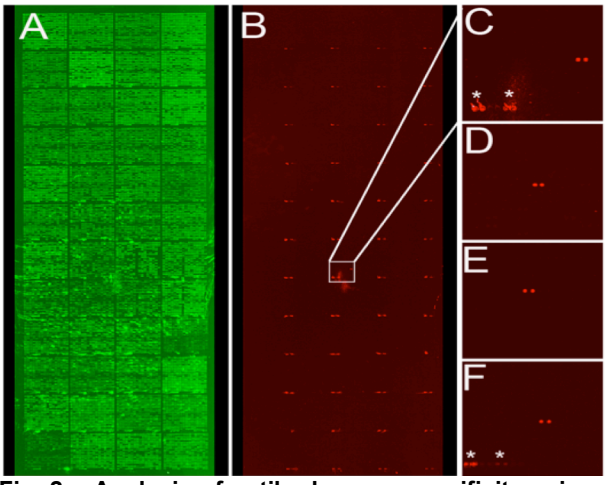

Fig. 2 – Analysis of antibody monospecificity using the Human Proteome Microarray (36)**.** Production of monoclonal antibodies (**mAbs**) against purified endogenous protein complexes (**Aim 2b**) will use the FastMab™ workflow and CDI's Human Proteome (**HuProt**) microarray (**Aim 2c**; **see *Approach***). **(A)** HuProt microarray probed with Cy3-coupled anti-GST antibodies, demonstrating the presence of recombinant protein. **(B)** HuProt probed with MAb recognizing HOOK1. **(C)** High-power image of box in B. Asterisks indicate Cy5-labeled IgG, which is printed four times in each protein block to act as a molecular beacon. Other examples shown include mAbs that selectively bind to TXDN9 **(D)**, WAS **(E)** and MTCP1 **(F)**. The HuProt microarray allows us to rapidly identify only the most promising candidate monoclonals to move forward for further validation in interactomic applications (**Aim 2c**) – establishing the cycle of production, validation, and new production (**see *Innovation***).

The principle challenge is to produce antibodies able to capture any chosen protein, while preserving any macromolecular complex that protein is part of, enabling analysis of the constituents as well as changes effected by differing physiological (e.g. disease) states. The most successful method currently available to explore the interactome, IP, most commonly implements affinity (or epitope) "tagging" of proteins of interest, permitting the tagged proteins and associated macromolecular partners to be purified from their crude biological source using a solid medium, frequently coupled to a common, commercially available antibody against the tag (affinity medium) (42,43). This is the only option when equivalent quality antibodies against the endogenous proteins are lacking. Although there are many benefits to using affinity tags, there are also numerous caveats arising from, e.g. altering the context of gene expression through to the possibility that the affinity tagging alters the behavior of the gene or protein (44); and tagging can only be carried out effectively in model organisms with efficient genomic tools, or which otherwise support ectopic gene expression, blocking the use of this strategy in many disease-relevant models and stymieing human interactome research. Antibodies against endogenous protein targets circumvent these issues because no alteration of the native gene or protein is required. However, producing antibodies against a wide-variety of targets (e.g. the human proteome) is time consuming and technically challenging, and not all antibodies are created equally.

Although there is a wide-variety of different kinds of antibody and antibody-like affinity reagents, the most widely utilized antibody reagents come in two flavors, polyclonal and monoclonal antibodies. Polyclonal antibodies can demonstrate excellent performance in affinity applications, but because these antibodies are an exhaustible resource and there are no guarantees that subsequent immunizations will produce a batch of equivalent performance, a race to failure is nevertheless initiated. Monoclonal antibodies (**mAbs**) are a renewable resource produced in cell culture by isolating and immortalizing individual (clonal) antibody producing B-cells harvested from immunized animals. Each mAb has a single kind of paratope but may still be of low affinity and specificity, binding numerous distinctive 'off-target' epitopes; or not be able to bind its target in the natural context of the macromolecular complexes it forms (45). Therefore effective strategies are required to identify high affinity, high specificity mAb clones which are able to IP endogenous protein

complexes: *this proposal includes independent strategies to (**Aim 1**) establish methods and validate CDI's extant collection of mAbs for their ability to IP endogenous protein* complexes as *well as to (**Aim 2**) efficiently produce monospecific mAbs against endogenous protein targets within macromolecular complexes* (**see *Innovation***).

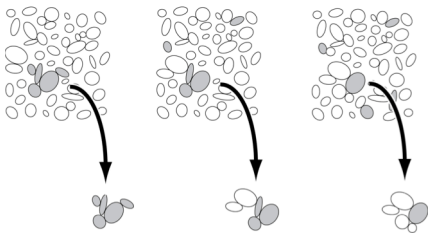

Fig. 3 – *Diagram illustrating the problems associated with disintegration of protein complexes once released from the cell.* Loss of cellular context, loss of authentic components (gray) and association of "false-positive" contaminants (white).

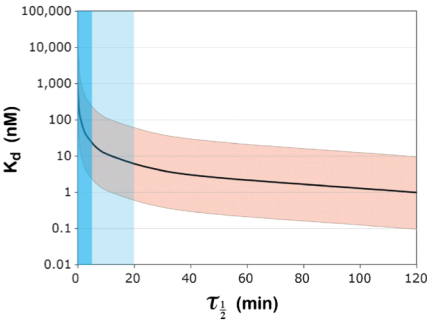

Fig. 4 – *Half-lives of binary complexes with the indicated K_ds and a K_{on} of 10^5 $M^{-1}s^{-1}$ (black line) and between 10^4–10^6 $M^{-1}s^{-1}$ (lower–upper orange limits).* Pale blue: Protein interactions exhibiting Kds in this range have post-extraction half-lives of 5–20 min. Dark blue: Protein interactions exhibiting Kds in this range have post-extraction half-lives of up to to 5 min. Most procedures last ~1 h or more, including extraction, centrifugation, batch binding, washing, and finally eluting. Optimizing the extractant to preserve the target complex can therefore benefit yield and fidelity enormously.

However, it is important to note that having an excellent antibody against a protein of interest is not enough to ensure success in protein complex purification. Even when high quality antibodies are available, in order to access and then isolate a protein and its associating partners, cells must be broken open and their contents released. In the absence of constant replenishment from a living cell, these macromolecular complexes will rapidly disperse. Dissociation of the complexes leads to intermingling of components not normally exposed to one another, and the resultant possibility of aberrant molecular interactions – a major source of "non-specific background" (**Fig. 3**). Thus, during isolation an artificial milieu consisting of buffers, salts, detergents, and other reagents, carefully selected in an effort to stabilize the complex, replaces the normal cellular context. Precise optimization of this artificial milieu (the extractant) is thus a crucial step that can have profound effects on the apparent affinities of the constituent interactions of a protein complex (1) (**Fig. 4**). A range of formulations to define a set of results that encompasses many distinct interaction states for the protein of interest must be explored, providing rich data on organization, function, and biochemistry. This remains an empirical science. Not surprisingly then, we are still far from a comprehensive and accurate map of the hierarchy of interactions within a cell, limiting our ability to effectively connect changes in the interactome with disease (4). As stated by John Bergeron, a former president of the Human Proteome Organization, "...many proteomics efforts suffer from a lack of rigor. Perhaps surprisingly, rather than acknowledging this, the proteomics community has instead attempted to circumvent the need for rigorous sample preparation by developing advanced bioinformatics tools. But this is only a very limited substitute for proper sample preparation, and …such efforts have had a limited impact in the cell biology community (15)." *We will (**Aims 1** and **2**) utilize a screening process (**Fig. 1**) that allows high throughput optimization of IPs, enabling rigorous selection of the extractant used for each individual macromolecular complex identified* (1) (**see *Innovation***).

From laboratory to marketplace: This Phase I STTR proposal seeks the support needed to improve mAb validation for use in interactomics and establish methods to vastly increase the coverage of mAbs against the human interactome in Phase II, eventually resulting in the commercialization of these reagents within the global biomedical communities. Although IP-based applications are the focus of this proposal, the resulting mAbs may also satisfy many other affinity-based applications in basic and translational research (41,45,46). Much successful research today is conducted by small-to-medium sized laboratories investigating a particular area: a segment of an interactome. A high level of expertise, interest, focus, and thoroughness are brought to bear on that area, but it is only possible to cover an entire system via the overlapping research of many such groups. The lack of antibodies to empower these laboratories to undertake in-depth investigations of the macromolecular interactions within their chosen area is a significant bottleneck, and is directly confronted by this proposal. Through the use of these antibodies to characterize specific changes in cancer PPIs, they are also likely to find immediate use in diagnostic and/or prognostic applications (6-8); while the interactomic understanding, if not the antibodies themselves, is likely to find therapeutic applications (11,12,30).

The research team: The synergy brought about by the current collaboration is significant. The desperate need for technologies that can quickly and reliably reveal the cellular interactome (47) is what drove the formation of the National Center for Dynamic Interactome Research (NCDIR, ncdir.org), of which Dr. LaCava is a senior member. This Center is devoted to the development of innovations to address the current shortcomings in interactomics analyses (48). Dr. LaCava has profound expertise in the development and practice of biochemical and proteomic methods (1,44,49-52); he is also a senior member of the Laboratory for Cellular and Structural Biology, at The Rockefeller University, which together with the

resources of the NCDIR, provides an incomparable level of research support. As one of two production centers over the past 4.5 years in the NIH's Protein Capture Reagent Program (http://proteincapture.org), CDI has optimized its mAb pipeline, which can now be applied to other large-scale antibody isolation projects.

Innovation – From two independent platforms (**Figs. 1 & 2**) we propose to generate a fused pipeline that permits cycles of mAb validation and interactome mining, followed – in a major innovation – by novel mAb generation using the discovered protein complexes as immunogens. *This cyclical system will expedite all interactomic analyses based on IP as well as expedite the production and characterization of new mAbs able to IP endogenous protein complexes*

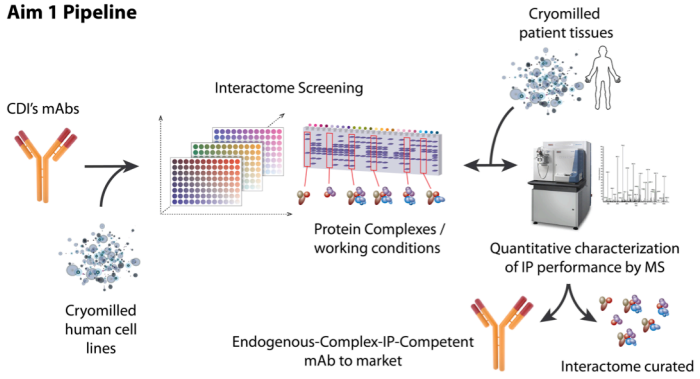

Fig. 5 (left to right) – mAbs drawn from CDI's anti-human TF collection will be subjected to IP optimization / interactome screening using cell lines (**Fig. 1**). Quantitative MS analyses will be used to describe the endogenous-complex-IP-competency of each mAb across a range or working conditions. mAbs will be subjected to comparable experiments utilizing patient-derived cancers. All results are curated at www.copurification.org for comparative analyses. Functionally robust antibodies are certified for IP and marketed commercially.

From an interactomic perspective there are two main innovations differentiating this proposal from other human interactome analyses. (I) All extant IP-based, large-scale human interactome studies have relied upon ectopic expression of affinity tagged protein in model cell lines as a means to capture complexes associated with each protein of interest; in contrast, **Aim 1** implements CDI's collection of unique mAbs, *circumventing tag- and expression-based artifacts and enabling IPs to be conducted on any source sample for each protein of interest* (**Fig. 5**). (II) Nearly all extant IP-based, large-scale human interactome studies have relied upon only a single condition of protein complex purification, sacrificing sample preparation quality for throughput (as described above in ***Significance***). Whereas our own system permits the exploration of up to ninety six conditions per protein of interest, per screen (many hundreds can be explored through repeat screening) (1) (**Fig. 1**). Each different condition utilized constitutes a novel attempt to capture the protein of interest in complex with its physiological interactors, providing dozens of alternative strategies per IP in *both* **Aims 1** and **2** (**Figs. 5 and 6**). When applied to a wide array of distinct protein IPs within a cellular pathway (such as TFs, see below), this system can improve conditions for the characterization and preparation of known and partially mapped complexes and discover new complexes, systematically exploring the interactome and revealing differences between healthy and diseased states (1,4).

From a biological perspective, Phase I will establish the methodological foundation to transform TF PPI analyses, and interactome analyses at large. With the working pipeline in place in Phase II we will explore the PPIs of a much larger number of TFs drawing upon the breadth of CDIs entire anti-TF mAb collection and we will be positioned to directly identify TF PPIs that are disease-linked as result of our analyses of patient cancers. Currently, the most common approaches to study TFs and their PPIs rely upon genomic and chromatin IP DNA-sequencing (ChIP-seq) data combined with computational prediction (26,53,54). Direct validations of these predictions are frequently lacking. Mass spectrometry (**MS**) based approaches that readout on protein-level associations have the potential to discover interactions that cannot be observed by e.g. ChIP-seq, which may reveal the DNA sequence preferences exhibited by TFs and putative interactors but cannot demonstrate PPIs directly. As detailed above in ***Significance***, TF interactomes remain very poorly characterized; alternative approaches that directly reveal PPIs are needed; our approach, fully described below, adopts a strategy to reveal TF PPIs with a direct IP-MS approach.

From the perspective of mAb production there are also two main innovations. (I) We shift away from the paradigm of mAb production using heterologously expressed protein immunogens, and thus away from the associated slew of caveats and provisos for recombinant human protein production in bacteria (55). Instead we will prepare endogenously assembled native protein complexes to use as immunogens (**Aim 2a, b; Fig. 6**). A sample handling 'trick' has permitted us to obtain endogenous complexes at the needed quantity and concentration to work as effective immunogens, ultimately yielding multiple mAbs to proteins in the mixture simultaneously (**Figs. 7** and **8**). (II) We will quantitatively assess of our mAbs' ability to IP endogenous protein complexes (**Aim 1b**) allowing us to ensure that we place functionally robust reagents into the market place providing consumers with reliable options that guarantee value for investment.

Approach – In Phase I we will independently test both proposed functions of the future (Phase II) fused pipeline. **Aim 1** focuses on establishing gold standards to characterize mAbs for use in interactomic analyses and implementing those mAbs to revolutionize the depth and accuracy of such analyses, revealing disease-specific features (**Fig. 5**). **Aim 2** focuses separately on the need to increase the rate of output of mAbs by using endogenously assembled affinity tagged multi-protein complexes as immunogens (**Fig. 6**). In Phase II, these two innovations will be fused together, *while also mitigating the reliance upon protein affinity tagging*, to further accelerate the cycle of mAb production and concomitant discovery.

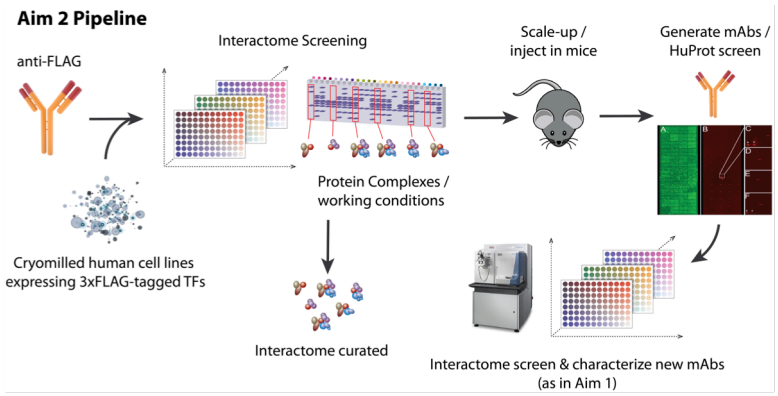

Fig. 6 (left to right) – 3xFLAG tagged TFs expressed in cell lines will be subjected to IP optimization / interactome screening (**Fig. 1**). Results will be curated on www.copurification.org for comparative analyses. Select conditions will be chosen for scale-up and entry into CDI's FastMab™ workflow (**Fig. 2**). Select monospecific mAbs will be entered into the Aim 1 pipeline (**Fig. 5**).

Preliminary studies: CDI's FastMab™ workflow leverages the highest content full-length human protein (**HuProt**) microarray (36), containing > 17,000 defined human proteins, to generate mouse mAbs that exhibit monospecificity for target human proteins (**Fig. 2,** further described below). To date, CDI have generated > 500 such mAbs against human TFs via this workflow. Each reagents has already demonstrated (I) monospecificity and (II) the ability to IP ectopically overexpressed target protein from HEK-293 cell extracts. Dr. LaCava's interactomic screen (1) permits the exploration up to ninety six unique protein complex purification conditions at a time, enabling thorough exploration of the parameters affecting success in IP-based interactomic analyses (summarized in **Fig. 1**).

The FastMab™ workflow: A 6-person production team is dedicated exclusively to the cell culturing steps involved in hybridoma production. They harvest immune cells from up to 24 mice, and perform an equal number of fusions, per week. *CDI is sufficiently equipped to support the additional requirements of a dedicated protein complex antibody production team* **(Aim 2).** The hybridoma fusion reactions are directly plated in semi-solid selective medium containing methylcellulose (MC), supporting growth of physically separated clones. Including fluorescently tagged anti-mouse IgG in the MC medium permits the selective collection of antibody (IgG)-secreting clones using microcapillary tubes, followed by clonal expansion. Antigen-binding mAbs are identified by exposing hybridoma culture supernatant to hybridization with immunogens spotted on mini-arrays. After mAb hybridization, a fluorescently tagged secondary antibody is used to permit detection of hybridization events in an array scanner. Immunogen binding mAbs are subsequently screened on the HuProt array to quantify their degree of monospecificity (36). A relative affinity (A score) is calculated using the normalized mean and standard deviation (SD) of the intensities of all spots on the array, and the specificity (S score) is the number of SDs of A score value between the highest A score ranking target and the second highest ranking target. To be considered monospecific a mAb must have an S score ≥ 3. Clones of interest are expanded, cells are frozen for permanent storage, and supernatant is collected for further analyses and ~10-50 mg-scale initial mAb production.

Specific Aim 1: Characterize CDI's anti-human TF mAbs for their ability to IP endogenous cancer-related TF protein complexes. To maximize the biomedical and commercial value of CDI's mAb reagents they must be characterized for their ability to IP their targets as part of naturally occurring protein complexes: 'endogenous-complex-IP-competency.' We will utilize the pipeline presented in **Fig. 5** to screen fifteen mAbs against six distinctive cancer-linked TFs (**Table 1**) for their ability to IP endogenous protein

Gene	Disease	# in PubMed	HEK-293 Abundance	HeLa Abundance	IHC type	IHC Abundance
NFKB1	Colonic Neoplasms	9	Low	Low	Colorectal, Ovarian	Low, Low
GATA4	Colorectal Neoplasms	2	Low	ND	Colorectal, Melanoma, Ovarian	High, High, High
SMAD4	Colorectal Neoplasms	15	Medium	Medium	Melanoma, Ovarian	Med, Med
SMAD7	Colorectal Neoplasms	4	Low	Low	Colorectal, Melanoma, Ovarian	Med, Low, Low
STAT3	Ovarian Neoplasms	8	Medium	Medium	Colorectal, Melanoma, Ovarian	Med, Low, Med
TP53	Colorectal & Colonic Neoplasms, Melanoma	355	High	Medium	Colorectal, Melanoma, Ovarian	High, Low, High

Table 1 – The TFs encoded by the listed gene (far right) will be initially targeted in **Aim 1** (**Fig. 5**); several attributes for each are listed. These targets were chose on the basis of: (I) extant mAbs in CDI's collection; (II) diverse targets, from well characterized highly expressed proteins to less well characterized lower expression proteins; and (III) classification of expression in skin, colon, and/or ovarian cancer. Disease associations and Pubmed data retrieved from www.disgenet.org. Abundance of the targets in HEK-293 and HeLa cell is based on RNA level; immunohistochemistry (IHC) data include types of patient cancers the listed TFs were detected in and their relative abundance as assessed by IHC; data available at www.proteinatlas.org.

complexes from HEK-293 and HeLa cells (**Aim 1a**). The first milestone will be to quantitatively characterize at least ten of these mAbs' endogenous-complex-IP-competency via MS: mAbs yielding the target and at least one bona fide interactor within the top five most abundant proteins will be categorized as endogenous-complex-IP-competent. The implementation (**Aim 1b**) we be comparable to recently established procedures (45), *but extended to characterize mAb function across multiple IP conditions*, which will be ranked according to the number and abundance of bona fide interactors (and independently validated by additional MS labs). If necessary to reach our milestones, we will draw additional mAbs from CDI's collection of over five hundred. Each mAb offers a unique opportunity to execute the entirety of **Aim 1**, the collection provides over five hundred alternative strategies. Utilizing the protocols and understanding gained by screening mAbs in HEK-293 and HeLa cells, endogenous-complex-IP-competent mAbs will be carried forward to (**Aim 1c**) *IP complexes from patient derived melanoma, colon, and ovarian cancers* (obtained from collaborators). The second milestone will be to validate up to five mAbs for endogenous-complex-IP-competency using patient derived cancers. We will also leverage the appropriate *NCI-60 cell lines as a complementary, alternative strategy*.

Specific Aim 2: Generate novel mAbs from purified endogenous TF protein complexes. To expedite mAb production we will adopt purified native human protein complexes comprising multiple proteins as immunogens. We will C-terminally 3xFLAG tag endogenous copies of the cancer-related TF genes listed in Table 1, all of which are expressed in HEK-293 cells. We will assess yield after single locus CRISPR-based tagging (56); tagging both loci if needed for low expression targets. We will utilize the pipeline depicted in **Fig. 6** to identify protein complexes copurifying with the targets and the conditions needed to obtain them (**Aim 2a**). Protein complexes containing components for which mAbs do not currently exist will be prioritized for scale-up and use as immunogens by CDI (**Aim 2b**). The first milestone will be to work up at least five such complexes for injection in mice. Our preliminary data has already demonstrated feasibility for this approach including the generation of sufficient quantities of complexes (**Fig. 7**) and the applicability of these complexes to the FastMab™ workflow (**Fig. 8**). Should additional TF targets be needed they will be selected based on high expression level in HEK-293 cells (www.proteinatlas.org), prioritized based on the number of cancer-related PubMed citations (www.disgenet.org), and tagged as described. Finally, newly generated mAbs will be deconvoluted using CDI's HuProt microarray (**Fig. 2**), and a selected set of mAbs with target S score ≥ 3 (monospecific) will be validated for endogenous-complex-IP-competency (as described in Aim 1; **Fig. 5**). The second milestone will be to generate up to fifty novel mAbs from TF protein complex immunizations.

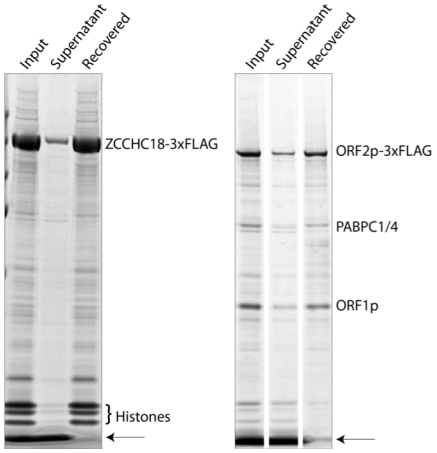

Fig. 7 – Concentrating Samples. Two distinct human protein complexes subjected to our procedure. Tagged and prominent copurifying proteins are labeled. After purification and native elution with 3xFLAG peptide, the complexes are concentrated under vacuum. As the volume decreases the 3xFLAG peptide remains dissolved (Supernatant) but the tagged protein complexes form a concentrated gelatinous pellet. The pellet is resuspended in 1/5th-1/10th the original volume of buffer (Recovered). Equivalent percentages of all samples were loaded. ≥ 80% of the input tagged protein complexes were recovered, concentrated by 5–10-fold, with near complete removal of 3xFLAG peptide (arrow). Multiple samples combined yielded ≳ 300 µg protein at ≳ 0.5 mg/ml concentration, which is suitable for immunizing mice and raising mAbs.

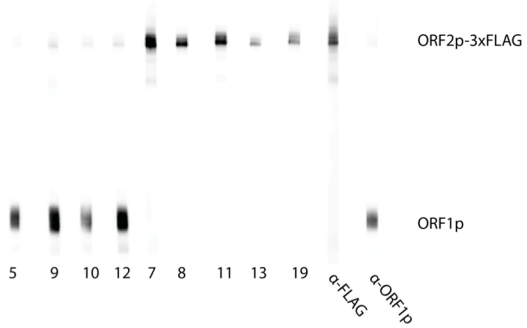

Fig. 8 mAbs From Complexes – ORF2p-3xFLAG protein complexes (LINE-1 retrotransposons (52)) from **Fig. 7** were subjected to FastMab™ workup. Twenty two hybridomas were obtained. These produced six mAbs with monospecificity to distinctive human proteins, assessed by HuProt array, as well as four mAbs against ORF1p and five mAbs against ORF2p-3xFLAG (both the most abundant proteins in the mixture; validated by western blot, shown above). In each lane purified ORF2p-3xFLAG complexes (the immunogen, **Fig. 7**) are loaded and each has been probed separately with purified mAbs from a numbered hybridoma clone (indicated). For comparison an anti-FLAG and an independently produced anti-ORF1p mAb are present, demonstrating that the CDI-generated mAbs bind the indicated target proteins. Clones 5, 9, 10, and 12 bind ORF1p, while clones 7, 8, 11, 13, 19 bind ORF2p. The anti-ORF2p mAb does not react with other FLAG-tagged proteins.

Summary – We will demonstrate feasibility to (I) create an iterative cycle of protein complex characterization and mAb generation and (II) lay the foundation to extend our approach to an interactome-wide scale. This will allowing us to vastly increasing our coverage of mAbs against the human proteome in Phase II, where the pipelines depicted in **Figs. 5** and **6** will be joined together as one in high throughput. The results of our analyses will enhance our ability to understand and directly address cancer through its interactome, starting with TF networks.