

Centromere round-up at the heterochromatin corral

Beth A. Sullivan

Studies in several organisms have shown that eukaryotic centromere regions contain proteins involved in chromosome inheritance functions such as centromere structure, heterochromatin assembly and sister chromatid cohesion. Mutation, inhibition, knockout and reciprocal epistasis experiments have revealed functions for many centromere region proteins and their order in the centromere assembly pathway. Proteins are organized into spatially distinct domains that might be established, separated and maintained by boundary elements and/or heterochromatin. Dissecting the multi-domain configuration of centromeres will enhance our knowledge of centromere and chromosome assembly and might aid in engineering artificial chromosomes for agricultural and gene therapy applications.

The centromere is a specialized chromosomal locus that ensures proper and equal segregation of chromosomes during cell division [1]. Many birth defects and cancers are caused by errors in chromosome inheritance. Therefore, understanding centromere organization and its various functions in chromosome inheritance has wide-ranging implications for research in human health, gene therapy and cancer drug development. In addition to testing sequences capable of supporting centromere formation, it is important to know all of the components and steps of assembly required for centromere formation. Several articles published in the past few months have discussed sequence composition and protein conservation in centromeres from yeast to mammals [2,3]. This article will focus on a few recent studies of centromere identity, organization and assembly in larger eukaryotes.

CENP-A, a conserved centromeric histone within kinetochore chromatin
Many centromere biologists have focused their research on the centromeric histone, CENP-A, a candidate for specification and propagation of kinetochore assembly [1,3]. CENP-A proteins are related to histone H3 [4]

and bind functional centromeric DNA (reviewed in [1] and [3]). Homologues have been identified in *Saccharomyces cerevisiae* (Cse4p), *Schizosaccharomyces* (Cnp1), *Caenorhabditis elegans* (HCP-3) and *Drosophila melanogaster* (CID), demonstrating evolutionary links amongst the widely divergent centromeric sequences of these organisms [1]. It has been suggested that sequence inconsistencies at centromeres might reflect adaptive mutation of CENP-A proteins to rapidly evolving centromeric DNA [5]. These models have yet to be tested directly and might be more applicable for evolutionarily distant or highly mutable genomes because epitope-tagged human CENP-A localizes to centromeres in distant mammalian species [6]. It is likely that CENP-A localization at centromeres involves a specialized, conserved secondary or tertiary structure.

Eukaryotic centromeres are organized into structural and functional domains
A particularly intriguing finding from cytological and molecular studies has been that centromeres are composed of many proteins arranged into physically separate domains. The centromeres of *S. pombe* are composed of non-overlapping protein domains: inner and outer repeats contain the proteins Swi6, Chp1, Rik1 and Clr4, whereas the central core contains Mis6 and Cnp1 (reviewed in [1] and [7]). Domain organization of the centromere region and the kinetochore is also present in multicellular eukaryotes [1,7]. Cytological studies in flies have demonstrated that CID/CENP-A chromatin is flanked by non-overlapping protein domains: inner and outer repeats contain the proteins Swi6, Chp1, Rik1 and Clr4, whereas the central core contains Mis6 and Cnp1 (reviewed in [1] and [7]). In addition, the centric cohesion protein MEI-S332 is located near, but not in, the CID-containing chromatin (Fig. 1), providing a physical basis for previous observations that kinetochore function and MEI-S332-mediated cohesion can be separated using minichromosome derivatives [9].

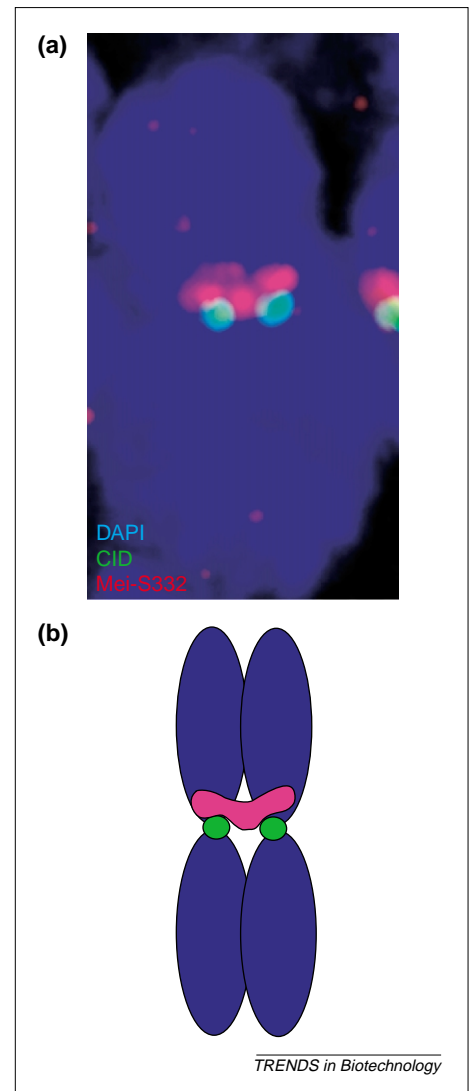


Fig. 1. Eukaryotic centromeres are organized into spatially distinct domains. (a) A metaphase chromosome from *Drosophila* S2 tissue culture cells stained using indirect immunofluorescence with antibodies to the kinetochore histone CID/CENP-A (green) and the centric cohesion protein Mei-S332 (shown in pink). CID and Mei-S332 staining do not co-localize. (b) Cartoon illustrating that the Mei-S332 (pink) cohesion domain does not overlap with the CID-containing kinetochore domain (green) and is situated to one side of the centromere. (Image courtesy of M. Blower.)

A general cohesion protein in humans, RAD21, occupies the inner pairing space on mammalian chromosomes and is located between CENP-B staining on sister chromatids of metaphase chromosomes [10]. Human RAD21 co-localizes with the primary

alpha-satellite array from chromosome 21, $\alpha 21$ -I, and extends into an adjacent, divergent alpha-satellite array, $\alpha 21$ -II, which is unable to support *de novo* kinetochore assembly [10,11]. Instead, $\alpha 21$ -II might function to establish centromeric cohesion. Functional centromeres on human artificial chromosomes (HACs) have been generated using synthetic or cloned alpha-satellite arrays that consisted of uniform higher order repeats, such as $\alpha 21$ -I. These exciting studies established that *de novo* mammalian chromosomes could indeed be engineered from component parts and suggested that these chromosomes could be developed for transferring genes into cells for gene therapy or for producing transgenic organisms. However, these first studies also indicated that all HACs that were generated were considerably larger than the original input DNA (~70–1000 kb) [11,12]. Size amplification might reflect requirements for multiple centromeric domains for proper chromosome assembly. It will be interesting to test if vectors containing centromeric DNA, as well as sequences associated with cohesion and heterochromatin proteins (i.e. a vector containing both $\alpha 21$ -I and $\alpha 21$ -II sequences), can efficiently generate *de novo*, stable chromosomes that are comparable in size to the original input DNA.

Functional relationships of centromeric domains

Mutation and reciprocal epistasis experiments in several species have revealed the functions of centromere region proteins and defined their positions in the centromere–kinetochore assembly pathway. Several lines of evidence indicate that the physical separation of some protein domains sometimes reflects functional independence. CID/CENP-A proteins appear to be near to or at the top of the hierarchy of factors that govern centromere–kinetochore formation and function. First, the localization of CID and the assembly and maintenance of centromeric chromatin is not affected in *Drosophila* mutants for *mei-S332* (sister chromatid cohesion), *prod* (condensation), *Su(var)2-5* (HP1) (heterochromatin) and *polo* (outer kinetochore) [8]. Second, inactivating CENP-A proteins in yeasts, worms, flies,

and mammals severely disrupts mitosis and cell cycle progression (reviewed in [1]). Furthermore, inactivating or deleting CENP-A causes mislocalization of many kinetochore and centromere-region proteins. CENP-A inactivation in mice and *C. elegans* abolishes the ability of CENP-C proteins to target to the kinetochore [13,14]. These effects do not extend to all neighbouring domains, because inhibition of *Drosophila* CID does not alter localization of the condensation protein PROD and heterochromatin protein HP1 [8].

‘These studies provide an excellent entry point for studying DNA–protein correlations and constructing maps of chromatin domains at human centromere regions.’

However, outer kinetochore proteins HCP1, BUB1 and MCAK are mislocalized when CENP-A is inhibited in worms [14,15] and outer kinetochore proteins Rod, Polo and CENP-E fail to localize when CID is inhibited in *Drosophila*. As a result, these cells show errors in chromosome segregation [8]. Similarly, MEI-S332 fails to localize in CID-depleted cells [8]. Therefore, cohesion domains containing MEI-S332 and RAD21–Scc1 are dependent on establishment of the CID/CENP-A domain, unlike HP1 and PROD, which are not epistatic to CID/CENP-A. These experiments in various species indicate that CENP-A localization in centromeric chromatin is independent of heterochromatin and condensation proteins and *vice versa*. However, outer kinetochore proteins require CENP-A for their localisation and for proper kinetochore assembly and chromosome behaviour.

Hierarchical assembly of centromere domains

Knockout and inhibition studies have also established the relationships of other centromeric proteins in the kinetochore assembly pathway. Disruptions of HCP-4 (CENP-C) in *C. elegans* and CENP-C and CENP-H in chicken tissue culture cells show no effect on CENP-A localization, but disrupted cells do fail to recruit outer kinetochore proteins, resulting in segregation defects and cell lethality [14–16]. CENP-H, a recently identified inner kinetochore protein, is required for CENP-C targeting

to centromeres, but not *vice versa* [16]. Thus, CENP-H and CENP-C are located downstream of CENP-A in the kinetochore assembly pathway. It is unclear if CENP-H directly binds to CENP-C, or if one or both proteins recognize a specific DNA sequence or chromatin structure. Nevertheless, these inner kinetochore proteins probably act closely with CENP-A in initiating assembly of the kinetochore and maintaining its structure. HCP-1, the *C. elegans* homologue of CENP-F, requires CENP-A and CENP-C for localization [15], indicating a functional role downstream of CENP-A. Finally, INCENPs, chromosomal passenger proteins that are involved in chromosome segregation and cytokinesis [17], are localized to the centromere region in metaphase, and INCENP–kinetochore association is required for delivering proteins to the metaphase plate or cleavage furrow. Neither INCENP nor CENP-A localization depend on each other in *C. elegans* or *Drosophila*, indicating that they make unique contributions to chromosome segregation [14,18].

The maxim of boundaries: ‘establish, preserve and protect’?

What factors ensure or maintain physical and functional distinctions between centromeric domains? In *S. pombe*, the inner-centromere and flanking-heterochromatin domains are separated by boundary elements, perhaps encoded by tRNA genes (Fig. 2a). Chromatin within the flanking heterochromatin domains contains H3 that has been methylated at lysine 9 (H3-K9), a modification that is necessary to recruit Swi6 and promote heterochromatin assembly and genome stability [19–21]. This histone modification is notably absent from the central core domain to which CENP-A (Cnp1) localizes. Another heterochromatic region, the mating type (*mat*) locus in *pombe*, also contains H3-K9 methylation and Swi6. This region also contains boundary elements; it is flanked by inverted repeats and the euchromatin outside of the inverted repeats contains H3 methylated at Lys4 (H3-K4), a modification that is notably absent within the centromeric domains [22]. On deletion of the repeats that flank the *mat* locus, H3-K9 methylation and Swi6 spread into neighbouring sequences and

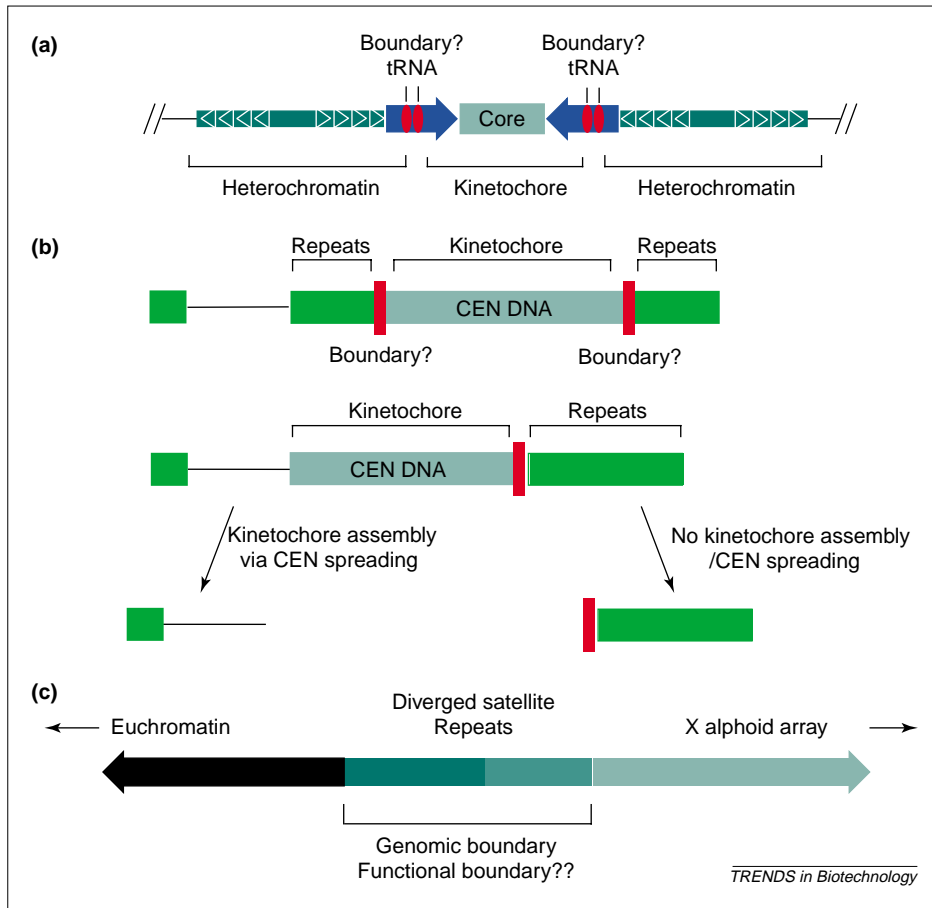


Fig. 2. Boundary elements at centromeric-heterochromatic domains might constrain or protect the transition from euchromatin to heterochromatin to kinetochore chromatin. (a) In *Saccharomyces pombe*, tRNA genes are putative boundary elements that flank the central core of the centromere to which CENP-A (Cnp1) localizes. H3 methylated at Lys9 is present in the inner and outer repeats but not within the central core. The tRNA genes may prevent either heterochromatin from spreading into kinetochore chromatin or vice versa. (b) In *Drosophila*, the centromere (CEN) of minichromosome *Dp1187* is flanked by heterochromatin (repeats). An inversion moves one side of the CEN from heterochromatin and places it next to euchromatin (solid black line). Irradiation-generated deletion derivatives of the inversion chromosome, *Dpγ238*, are recovered that lack CEN DNA but contain CID/CENP-A and other kinetochore proteins, suggesting that only centromeric chromatin that has lost a heterochromatin boundary can spread into neighbouring euchromatin. Derivatives lacking CEN DNA are not generated from region of *Dpγ238* that retains heterochromatic boundaries. (c) A newly identified genomic boundary between the border of the human X centromere and short arm euchromatin extends for ~450 kb and is composed of divergent satellite sequences. Further studies will determine if this genomic boundary correlates with a functional boundary.

inhibit H3-K4 methylation.

These findings suggest that distinct transition zones restrict or protect heterochromatin-euchromatin domains, separating unique patterns of histone modifications that denote active, silent, or specialized chromatin [5,22]. Chromosomes in larger eukaryotes possess boundary elements that define imprinted domains; therefore, centromeric regions are also likely to be structurally regulated by heterochromatin, which may 'corral' kinetochore chromatin and prevent it from spreading into adjacent domains [23]. Evidence for heterochromatin performing this function comes from studies in which heritable *Drosophila* neocentromeres were induced experimentally only when test

fragments were juxtaposed to an active centromere but not when the test segments were located further from centromeres (Fig. 2b) [23]. CID was detected on these neocentromeres [8], suggesting that in *Drosophila*, centromeric chromatin that is stripped of heterochromatin boundaries can spread to adjacent DNA, similar to observations in yeast [22,24]. In humans, a genomic boundary spanning the border of the primary alpha-satellite array of the centromere and the first expressed sequences within euchromatin on the X chromosome has been only recently described [25] (Fig. 2c). This transition region contains 450 kb of diverged satellite repeats. These studies provide

an excellent entry point for studying DNA-protein correlations and constructing maps of chromatin domains at human centromere regions.

Centromeric chromatin is spatially separable from canonical centric heterochromatin and there are distinct spatial and functional domains within the centromere and adjacent regions of eukaryotes. These domains are characterized by specific proteins and might be defined and maintained by heterochromatin and chromatin modifications. Inhibition and epistasis analyses suggest that CID/CENP-A proteins are near or at the top of the hierarchy of factors governing centromere identity, propagation, and kinetochore formation. Kinetochore, condensation and chromosomal passenger proteins occupy different structural and functional domains that are required for mitotic chromosome structure and segregation. Identifying proteins that interact with or load CENP-A onto chromatin, determining how CENP-A proteins interact with diverse centromeric DNAs, and defining protein-DNA interactions within centromeric domains are important next steps for advancing our knowledge of centromere and chromosome biology. In addition, isolating the components required for centromere formation and function will facilitate further development of artificial chromosomes for applications in biomedicine and biotechnology.

Acknowledgements

I thank M. Blower and H. Le for the image and cartoon in Fig. 1, J. Bridger, M. Blower, H. Le and A. Skora for comments and suggestions, and K. Sullivan and G. Karpen for support and encouragement. I apologize to those colleagues whose work I could not directly reference owing to space constraints.

References

- 1 Sullivan, B.A. *et al.* (2001) Determining centromere identity: cyclical stories and forking paths. *Nat. Rev. Genet.* 2, 584-596
- 2 Kitagawa, K. and Hieter, P. (2001) Evolutionary conservation between budding yeast and human kinetochores. *Nat. Rev. Mol. Cell. Biol.* 2, 678-687
- 3 Sullivan, K.F. (2001) A solid foundation: functional specialization of centromeric chromatin. *Curr. Opin. Genet. Dev.* 11, 182-188
- 4 Palmer, D.K. *et al.* (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3734-3738

- 5 Henikoff, S. *et al.* (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293, 1098–1102
- 6 Sullivan, K.F. *et al.* (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* 127, 581–592
- 7 Choo, K.H. (2001) Domain organization at the centromere and neocentromere. *Dev. Cell* 1, 165–177
- 8 Blower, M.D. and Karpen, G.H. (2001) The role of *Drosophila* CENP-A/CID in kinetochore formation, cell-cycle progression and interactions with heterochromatin. *Nat. Cell Biol.* 3, 730–739
- 9 Lopez, J.M. *et al.* (2000) Sister-chromatid cohesion via MEI-S332 and kinetochore assembly are separable functions of the *Drosophila* centromere. *Curr. Biol.* 10, 997–1000
- 10 Hoque, M.T. and Ishikawa, F. (2000) Human chromatid cohesin component hRad21 is phosphorylated in M phase and associated with metaphase centromeres. *J. Biol. Chem.* 16, 5059–5067
- 11 Ikeno, M. *et al.* (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat. Biotechnol.* 16, 431–439
- 12 Harrington, J.J. *et al.* (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nat. Genet.* 15, 345–355
- 13 Howman, E.V. *et al.* (2000) Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1148–1153
- 14 Oegema, K. *et al.* (2001) Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. *J. Cell Biol.* 153, 1209–1226
- 15 Moore, L.L. and Roth, M.B. (2001) HCP-4, a CENP-C-like protein in *Caenorhabditis elegans*, is required for resolution of sister centromeres. *J. Cell Biol.* 153, 1199–1208
- 16 Fukagawa, T. *et al.* (2001) CENP-H, a constitutive centromere component, is required for centromere targeting of CENP-C in vertebrate cells. *EMBO J.* 20, 4603–4617
- 17 Mackay, A.M. and Earnshaw, W.C. (1993) The INCENPs: structural and functional analysis of a family of chromosome passenger proteins. *Cold Spring Harbor Symp. Quant. Biol.* 58, 697–706
- 18 Adams, R.R. *et al.* (2001) Essential roles of *Drosophila* inner centromere protein (INCENP) and aurora B in histone H3 phosphorylation, metaphase chromosome alignment, kinetochore disjunction, and chromosome segregation. *J. Cell Biol.* 153, 865–880
- 19 Nakayama, J-I. *et al.* (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292, 110–113
- 20 Bannister, A.J. *et al.* (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120–124
- 21 Peters, A.H.F.M. *et al.* (2001) Loss of the *Suv39h* histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107, 323–337
- 22 Noma, K. *et al.* (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* 293, 1150–1155
- 23 Maggert, K.A. and Karpen, G.H. (2001) The activation of a neocentromere in *Drosophila* requires proximity to an endogenous centromere. *Genetics* 158, 1615–1628
- 24 Partridge, J.F. *et al.* (2000) Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev.* 14, 783–791
- 25 Schueler, M.G. *et al.* (2001) Genomic and genetic definition of a functional human centromere. *Science* 294, 109–115

Beth A. Sullivan

Molecular and Cell Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037 USA.
e-mail: bsullivan@salk.edu

Gene targeting comes to top-down drug screens

John M. Sedivy

With very few exceptions, cancer drug screening has been stuck with the multiple variable approach. While those of us working with model genetic systems would regard with horror experiments using non-isogenic strains or outbred lines, cells that differ from common human tumor cells by simple and/or defined genetic traits are generally not available. A recent paper published by Torrance *et al.* takes an important step towards a more rational and controlled approach in the area of cancer drug screening.

Most of us have had the experience of asking a student or coworker what they did to make the experiment finally work and being answered with a laundry list of modifications. More often than not the further query: 'and which one made the crucial difference?' will elicit the answer: 'I was desperate so I tried everything I could think of at once.' Although some of us are gifted with sufficient intuition to succeed in this manner, it is subsequently necessary to go back and test individual variables until the causative agent is found. For the less precocious (or lucky) the tried and tested method is to vary one parameter at a time.

The paper by the Kinzler and Vogelstein group [1] takes an important step towards a more rational and controlled approach towards cancer drug screening. They started with a human colon cancer cell line (DLD-1) in which the Shirasawa group had previously disrupted an activated *Ki-Ras* R13W allele using gene targeting [2]. They went on to stably modify the *Ki-Ras* +/- cell line with blue fluorescent protein (BFP) and its parental *Ki-Ras* +/(R13W) cell line with yellow fluorescent protein (YFP). This allowed the co-culture of the two cell lines and a real-time, non-invasive assessment of the relative composition of the cultures using simple spectral analysis. Thus, if a drug were added that differentially affected one or the other cell line, the co-culture would change color. This clever method provided not only an internal control but also lent itself easily to scale-up for a high throughput drug screen.

It is important to note that Torrance *et al.* chose ablation of an endogenously activated *Ras* allele rather than introducing a mutant *Ras* gene into a cell that previously lacked it.

Although conceptually similar, the latter approach can be confounded by the fact that excessive expression of mutant *Ras* genes can trigger growth inhibitory effects [3]. Thus, although engineering gain-of-function mutations is an important and powerful tool, investigators must be mindful of maintaining physiological levels as well as patterns of expression. In this regard 'knock-in' gene targeting interventions would be the ideal way to proceed.

'The paper... takes an important step towards a more rational and controlled approach towards cancer drug screening.'

Careful attention to such details should enable the eventual 'construction' of human tumor cells from normal cells [4] by recreating a clinically relevant series of genetic events. Multiple genetic interventions, either to reconstruct or deconstruct a cancer cell, would also be desirable in studies aimed at discovering drugs that impair one or more precursor cell types without affecting the ultimate normal precursor. Although this is conceptually similar to current methods of