Genetic modification for cell pedigree labels to aid disease treatment

Ronald L. Rivest
rivest@mit.edu

March 10, 2014

Abstract

We suggest modifying the human genome in order to aid in the treatment of disease. The modifications enable targeted treatment of all cells descended from a given cell, by providing each cell with a "*cell pedigree label*" (CPL). We specifically consider the treatment of cancer.

1 Introduction

What modifications to the human genome would aid in the treatment of diseases such as cancer?

We suggest genetic modifications that would provide a *label* to each cell, so that cells with identical, similar, or related labels could be selectively treated. Cells selected for treatment could be killed or subjected to a disease-specific treatment. Cells not selected would be unaffected.

Our proposal envisions that the genetic modification is present from conception. A fertilized egg would contain the genetic modifications, as would all of its descendants. The modifications ensure that every cell is provided with a label. The labels reflect the "phylogenetic" tree of mitotic divisions that created the body from the original fertilized egg. The label of a cell reflects its lineage or pedigree, so we call these labels "*cell pedigree labels*" (CPLs).

We propose that the genetic modifications and labels have the following properties:

Coverage: Every cell contains at least one label.

Variability: Different cells may contain different labels.

Observability: Given a cell, it is possible to determine its label(s).

Changeability: The label(s) of a cell may change over time.

Inheritance: The initial label(s) of a daughter cell is derived from the current label(s) of its parent, but may be different than the current label(s) of its parent.

Computable ancestor labels: From the labels of a set of cells one can compute the label(s) of their most recent common ancestor (MRCA) (or last common ancestor)—the last or most recent cell that is an ancestor of all of the given cells [1]

Selective treatment: One can selectively treat those cells (and only those cells) descended from a cell having a given label or (set of labels).

One can use these properties to kill a tumor: extract a number of tumor cells and read their labels. Compute the label of their most recent common ancestor. Treat and kill all cells descended from that most recent common ancestor. This approach works well, even if the cancer has metasisized and spread throughout the body, since we are relying on inherited characteristics, not physical location, to identify the cancer cells.

A fertilized egg may contain a short label, unrelated to the labels of the germ cells that created it.

[1] This terminology is usually applied to individual organisms rather than cells, but its usage here is natural.

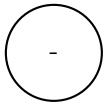

Figure 1: The initial cell (e.g. a fertilized egg) with no label.

2 Implementing Labels

We imagine that a label is represented as a sequence of letters from some alphabet, or a set of such sequences. Our examples use the alphabet `X,Y`. (These should *not* be confused with the use of X, Y to denote gender-specific chromosomes!)

A label may be encoded and represented within the nuclear DNA of the cell. For example, one might represent `X` as `AGTCATGAACA` and `Y` as `GGACTGCATT`. The representation of each letter from the label alphabet would be as a sequence of DNA base pairs that does not interact with or interfere with the other normal operations of DNA within the cell.

Here are three ways to implement labels.

Strict tree labels Here the root of the tree (the fertilized egg) has some arbitrary label L_0, such as the empty label.

If a cell with label L undergoes mitosis, its two daughter cells have labels L`X` and L`Y`. The labels of the daughter cells are one letter longer than the label of the parent. Every cell gets a unique label.

The length of a cell's label, minus the length of L_0, gives the "generation number" of the cell.

Given a set of labels, the longest string that is a prefix of all of them is the label of their most recent common ancestor.

Relaxed tree labels These are like strict tree labels, but the operation of extending the label of a cell is decoupled from mitosis.

In this model, a daughter cell inherits the label of its parent.

Furthermore, a cell may change its label at any time, by randomly adding either an `X` or a `Y` to the end of its label.

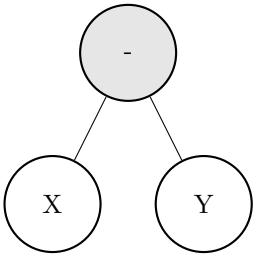

Figure 2: First mitosis. Time proceeds from top to bottom. The initial cell, with no label, splits into two daughter cells, labeled "X" and "Y". The initial cell is colored gray as it is no longer present. Its daughter cells are white as they are alive and presently part of the living organism.

The labels are randomly growing strings; the rate of growth of the string may be determined by factors other than the mitosis rate. Perhaps each cell adds several random letters to its label before it undergoes mitosis itself.

Again, given a set of labels, the longest string that is a prefix of all of them is a label of their most recent common ancestor.

(Strict or Relaxed) Set labels Here, a label consists of a set of strings over the label alphabet.

When mitosis happens (strict) or at other times (relaxed), the set may be augmented by the addition of another string. This new string may be randomly chosen according to some distribution.

Given the label for a set of cells, a label for their most recent common ancestor is the set intersection of the labels of the cells.

3 Observability and Selective Treatment

My brother (who, unlike me, is a biologist!) suggested that one might both obtain observability of labels and enable selective treatment by having labels expressed as proteins (antigens) that embed them-

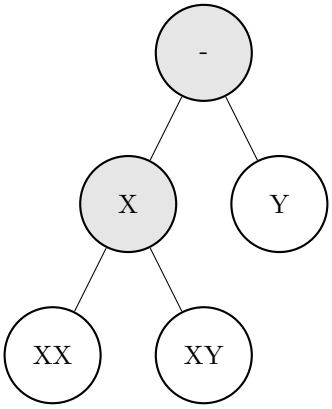

Figure 3: Second mitosis. The cell with label "X" splits into two daughter cells, labeled "XX" and "XY". Cell "X" is colored gray as it is no longer present. Its daughter cells are white as they are present. Note that the label of a daughter cell always extends the label of its parent.

selves stably in the cell's surface and become observable biomarkers.

Immunotherapy using monoclonal antibodies might then become useful in killing the selected cells. For example, Lampson [1, 2] discusses the successes and challenges of using monoclonal antibodies to treat brain tumors, based on existing markers for the cancer cells.

4 Discussion

The proposed method is well-tuned to fighting cancer, since cancer cells are typically all descended from a single mutant ancestor cell.

The proposed method is also suitable for treating other diseases, when it is desired to target a specific organ or portion of an organ that is derived from a common ancestor cell.

There are many reasonable variations on the proposed method. For example, cell might might have more than one label—a cell could have its own label but also the labels of its ancestors. Labels might propagate from generation to generation, instead of starting fresh with each generation. Label extension could be triggered by external stimuli. Labels may contain "error-correction" so that mutations within the labelling mechanism itself are less likely to cause problems with the use of this method. The entire labeling mechanism might be duplicated two or more times in parallel, with distinct sets of markers, so that the failure of one mechanism might still leave another one available for use. (See Lampson [1] for discussion of the need for multiple attack mechanisms in attacking brain tumors.)

The Hayflick limit suggests that the number of times a cell undergoes mitotic division is at most 40 to 60. The length of a cell label may thus be roughly 40 to 60 letters long. (Cancer cells, however, may ignore this "limit.")

Cell labels allow the capture and reconstruction of ontogeny, just as DNA sequencing and analysis allow the reconstruction of the tree of evolution. The computational problem here is much simpler, since variability is not introduced by random mutations, but rather by carefully designed mechanisms.

The method proposed here is not a way to combat cancer in the current human population, but is rather an idea that might enable future generations to fight cancer more effectively.

There is *much* work to be done to see if the proposed method can be made to work. Making the labels observable, making them appropriately variable,

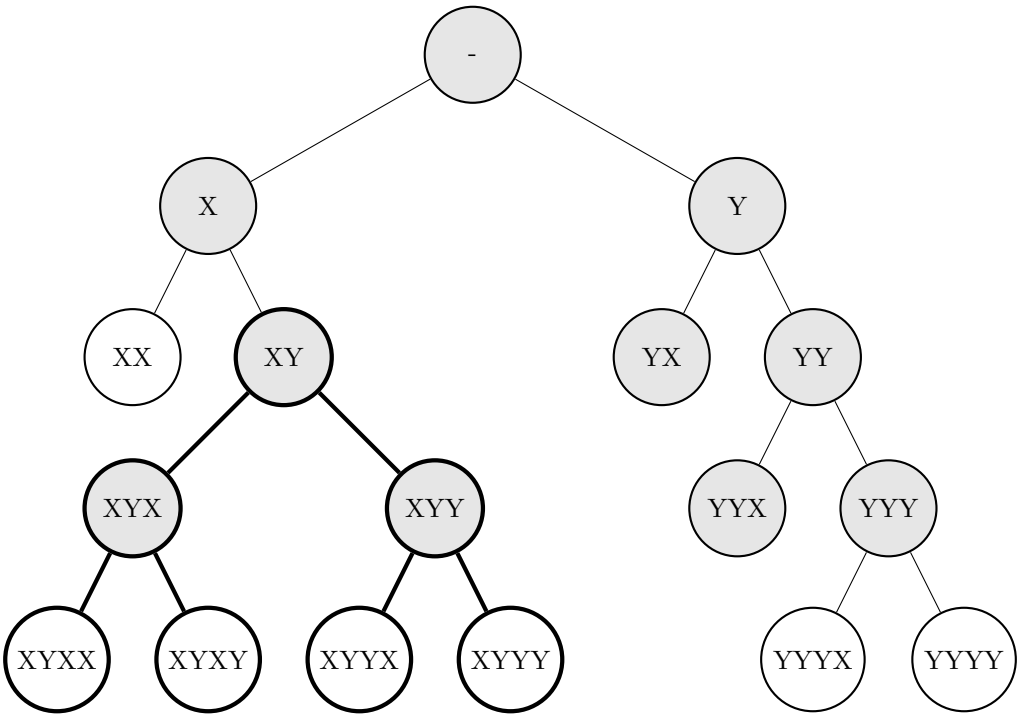

Figure 4: The phylogenetic tree after eight splits. Seven cells are alive (white), while ten are no longer present (gray). Note that cells "YX" and "YYX" died without producing daughter cells. Suppose that cell "XY" had a mutation leading to a cancer. Its descendant cells are easily identified as those whose label begins with "XY". The subtree of descendants of "XY" is darkened for emphasis. Killing all four living cells whose label begins with "XY" will completely remove the cancer, and nothing else.

and making treatment appropriately selective will no doubt be difficult. There does not seem to be any essential reason why these properties should not be achievable. Synthetic biology and genetic engineering are making great progress.

Eric Lander (private conversation) suggests that the high mutation rate of cancer cells may already provide suitable CPL's, so that genetic engineering may be unnecessary.

There is obvious value to having such a cell-labelling method, even if selective treatment is not possible with that labelling method. It would be very interesting to "read out" the morphogenetic structure of an organism from the labels on its cells.

5 Conclusion

We have sketched a way in which genetic modifications could aid in the treatment of diseases such as cancer. The idea is simple but powerful. The real challenge is making such an idea work!

Acknowledgment

The author gratefully acknowledges support from the MIT Vannevar Bush Professorship. The author also thanks Steven Averbuch, Tom Knight, Lois A. Lampson, Eric Lander, Alex Rivest, and Brian Rivest for helpful comments. Any errors or omissions are my fault, however, not theirs.

References

[1] Lois A. Lampson. Targeted therapy for neuro-oncology: reviewing the menu. *Drug Discovery Today*, 14(3/4):185–191, February 2009.

[2] Lois A. Lampson. Monoclonal antibodies in neuro-oncology: getting past the brain-blood barrier. *mAbs*, 3(2):153–160, March/April 2011.