

identifiers (“anonymizing” the sample), the possibility of risk to the subject stemming from the results of the research is greatly reduced. Large-scale DNA sequence determination represents an exception because each person’s DNA sequence is unique and, ultimately, there is enough information in any individual’s DNA sequence to absolutely identify her/him. However, the technology that would allow the unambiguous identification of an individual from his/her DNA sequence is not yet mature. Thus, for the foreseeable future, establishing effective confidentiality, rather than relying on anonymity, will be a very useful approach to protecting donors.

Investigators should introduce as many disconnects between the identity of donors and the publicly available information and materials as possible. There should not be any way for anyone to establish that a specific DNA sequence came from a particular individual, other than resampling an individual’s DNA and comparing it to the sequence information in the public database. In particular, no phenotypic or demographic information about donors should be linked to the DNA to be sequenced.² For the purposes of the HGP such information will rarely be useful, and recording such information could result in possible misuse and compromise donor confidentiality.

Confidentiality should be “two way.” Not only should others be unable to link a DNA sequence to a particular individual, but no individual who donates DNA should be able to confirm directly that a particular DNA sequence was obtained from their DNA sample.³ This degree of confidentiality will preclude the possibility of re-contacting DNA donors, providing another degree of protection for them. It should be clear to both investigators and to donors that the contact involved in obtaining the initial specimen will be the only contact.⁴

Another approach for protecting all DNA donors is to reduce the incentive for wanting to know the identities of particular donors. If the initial human sequence is a “mosaic” or “patchwork” of sequenced regions derived from a number of different individuals, rather than that of a single individual, there would be considerably less interest in who the specific donors were. Although there may be scientific justification that each clone library used for sequencing should be derived from one person, there is no scientific reason that the entire initial human DNA sequence should be that of a single individual. As approximately 99.9% of the human DNA sequence is common between any two individuals, most of the fundamental biological information contained in the human DNA sequence is common to all people.

To increase the likelihood that the first human DNA sequence will be an amalgam of regions sequenced from different sources, a number of clone libraries must be made available. Although a number of large insert libraries have been made,

most do not meet all of the standards set in this document; therefore, these libraries should be used as substrates for large-scale sequencing only under circumscribed conditions (see section 6, p. 79). Starting immediately, new libraries will be developed that have the advantage of being constructed in accordance with the ethical principles discussed in this document; they may also confer some additional scientific benefit. Such libraries are critical for the long-range needs of the HGP.

3. Source/Recruitment of DNA Donors for Library Construction

Another implication of the fact that 99.9% of the human DNA sequence is shared by any two individuals is that the backgrounds of the individuals who donate DNA for the first human sequence will make no scientific difference in terms of the usefulness and applicability of the information that results from sequencing the human genome. At the same time, there will undoubtedly be some sensitivity about the choice of DNA sources. There are no scientific reasons why DNA donors should not be selected from diverse pools of potential donors.⁵

There are two additional issues that have arisen in considering donor selection. These warrant particular discussion:

- It is recognized that women have historically been underrepresented in research, so it can be anticipated that concerns might arise if males (sperm DNA) were used exclusively as the source of DNA for large-scale sequencing. Although there would be no scientific basis for concern, because even in the case of a male source, half of the donor’s DNA would have come from his mother and half from his father, nevertheless perceptions are not to be dismissed. While the choice of donors will not be dictated to investigators, it is expected that, because multiple libraries will be produced, a number of them will be made from female sources while others will be made from male sources.
- Staff of laboratories involved in library construction and DNA sequencing may be eager to volunteer to be donors because of their interest and belief in the HGP. However, proximity to the research may create some special vulnerabilities for laboratory staff members. It is also possible that they will feel pressure to donate and there may be an increased likelihood that confidentiality would be breached. Finally, there is a potential that the choice of persons so closely involved in the research may be interpreted as elitist. For all of these reasons, it is recommended that donors should not be recruited from laboratory staff, including the principal investigator.