

Commentary

Nature Reports Stem Cells

Published online: 10 July 2008 | doi:10.1038/stemcells.2008.85

How to assess a stem cell genome

Michael A Teitell¹

To pick the best line for the job, genetic variation should be better understood



Michael A Teitell

As potential therapies derived from pluripotent stem-cell lines move towards the clinic, researchers have yet to address an essential, practical question: how should cells be screened for genetic variants that contribute to cancer or other diseases? The tools to address this question exist, but discussions on how to use them for this purpose are rare and uncoordinated.

Stem cell biologists are understandably more obsessed with epigenetics than with genetics. In a developing human embryo, epigenetic controls over gene expression coordinate the creation of several hundred cell types to generate an entire organism. Remarkably, laboratory techniques can trigger epigenetic reprogramming of somatic cells to an embryonic-like state, and scientists are learning how to direct such cells into the types useful for research and therapies. Improvements in techniques (in both nuclear transfer and direct reprogramming) could soon mean that more pluripotent cell lines can be generated for particular purposes or even for particular patients. That is precisely why the field should give more attention to genetic variation within stem cell lines.

Embryonic stem (ES) cell lines vary dramatically in their proclivity to differentiate into particular lineages¹⁻². Although some of this variation will be related to chance events and culture techniques, some differences will certainly come down to the genetics of human populations. Understanding this contribution will become even more important as more cell lines are derived from individuals with specific disease susceptibilities.

Much human variation can be pinned on differences in DNA. Genome-wide association studies identify loci with small effects that influence a person's chance of getting a range of diseases, including inflammatory diseases and cancer. Gene-linkage analyses, such as the linking of mutations in *BRCA1* and *BRCA2* genes to breast cancer, have found a smaller number of loci with larger effects. The lists of dangerous sequences grow with every upload to PubMed.

The recent examination of James Watson's genome provides a good thought experiment. Massively parallel DNA sequencing assessed more than 3.3 million single nucleotide polymorphisms (SNPs) in Watson; these were then compared with a human 'reference genome'³. That analysis revealed several variants linked to illness and about 11,000 SNPs affecting the amino-acid sequence of various proteins. If a pluripotent stem cell line were made from Watson's cells, how could scientists use genetic analyses to determine whether this line would be more or less suitable than others for a particular study or therapy? What established disease-susceptibility genes should they test for?

In short, it is time to consider how to assess stem cell genomes, and to discuss whether genomes associated with disease predisposition should be excluded from the therapeutic arsenal.

It is time to consider how to assess stem cell genomes and to discuss whether those associated with disease predisposition should be excluded from the therapeutic arsenal

Counting chromosomes is not enough

Right now, the one quality that pluripotent stem cell genomes are routinely assessed for is stability: do cells maintain the normal number and gross configuration of chromosomes over long periods in culture? This overlooks finer genome rearrangements, including ones that could, presumably, make cells dangerous in the clinic and unreliable in the laboratory.

At the very least, it seems, stem cell lines should be screened for genetic variants predisposing cells to uncontrolled growth, but even for that defined purpose, setting the level of rigour is difficult. Insightful data-mining techniques recently implicated the unsuspected *TMPRSS2-ERG* fusion gene in prostate cancer⁴. At what point should an institution hoping to use a cell line for therapy screen lines for such oncogenic variants? What if a susceptibility locus for heart disease was in a cell line being prepared for transplantation into children with heart disease? Those transplants could outlast the scientists and regulators who green light the clinical trial. At what point in clinical development should these questions be raised? By the time cell lines carrying suspected variations are ready to move forward, such variants might be recognized as either harmless or unacceptable.

Stem cells have multiple sources of genetic variation

Assessing genome quality will also mean understanding the sources of variation. Certain genetic changes are likely to arise only at distinct stages of human ES cell generation, maintenance and differentiation. The long-term effects of intentional genomic modifications, such as inserting reporter genes, are also largely unassessed. In other words, assessing genome quality means not just asking what the variants are, but also when that variation occurs.

A crucial distinction is whether variations have arisen before or after material taken from individuals is used to generate a cell line; for an ES cell line, variations not found in either gamete donor could arise during gametogenesis, embryo development or manipulation. Extended cell culture may select for other variants. It is well established that some gross chromosomal changes tend to accumulate over culture time⁵ and smaller, currently undetected changes may accumulate as well.

A basic assessment would compare genomes in the starting materials and in established cell lines. For human ES cell lines created from left-over embryos this could be problematic. Even if tissue from the parents was available, it could be unreliable. Chromosomes within gametes are different from those in progenitor cells; for these and other reasons, genomes in tissue collected from adults may not represent the genomic contribution to an embryo.

For somatic cell genomes reset to pluripotency, sufficient material for analysis can probably be gathered from the initial biopsy and the established pluripotent stem cell line. Such analyses

would pick up any variation between the patient's genome and an established stem cell line. Arguably, if cell lines are generated from specific patients, stability will become the most important criterion, as the patient and the cells will have the same genome. Even when human pluripotent stem cell lines are generated without viruses or through nuclear transfer, the act of reprogramming and culturing might still introduce selective pressures that favour certain genetic alterations over others.

Many techniques are available

Currently, genome integrity for both human ES cells and induced pluripotent stem cells (iPS cells) is mainly determined by classical G-banding cytogenetics — the Giemsa staining of metaphase chromosomes, which are compact enough to be visible in the light microscope. Very occasionally, a technique called metaphase-based comparative genomic hybridization (CGH) is used to assess human ES cells. These techniques detect gross chromosomal aberrations such as aneuploidy (the lack or extra copy of a chromosome). But G-banding-based karyotyping can only detect features 5–10 Mb in size and CGH can detect those of 2–3 Mb, so genomic changes of less than 2 Mb go unnoticed. None of the thousands of variations of the sort identified in Watson's genome can be considered, nor can differences in how many copies of a gene exist within a genome; such copy number variant (CNV) regions are increasingly recognized as important.

Technique	Examine	Detect	Limitations
Cytogenetics (G-banding is most common)	Giemsa-stained whole chromosomes	Aneuploidy Amplifications Deletions Rearrangements and translocations	Low resolution (5–10 Mb)
Comparative genomic hybridization (CGH)	Two-color fluorescence-stained whole chromosomes	Aneuploidy Amplifications Deletions	Low resolution (2–3 Mb) Requires reference genome
Spectral karyotyping (SKY) Multiplex fluorescence <i>in situ</i> hybridization (M-FISH)	Multicolor fluorescence-stained whole chromosomes	Aneuploidy Rearrangements and translocations	Low resolution (1–5 Mb)
Array CGH (aCGH)	Hybridization to microarrays containing cloned DNAs, cDNAs, or oligonucleotides	Aneuploidy CNVs (copy number variants, that is, deletions and amplifications) SNPs (single nucleotide polymorphisms)	Medium resolution (10–100 kb) Requires reference genome Locations of amplifications and rearrangements unknown Some designs do not distinguish between parental genomes
Massively parallel sequencing (1 G Illumina Solexa; Roche 454; paired-end sequencing)	nucleotide order (20 bp to 400 bp or more, depending on system)	DNA sequences	Can be expensive Requires bioinformatics skills Requires reference genome
Of the techniques listed above, only G-banding cytogenetics is used routinely.			

The genetic variability known to influence human traits has not yet been assessed in pluripotent stem cell lines, but there are several approaches for doing so (Table 1). Several relatively rapid, massively parallel genome-wide sequencing methodologies are available^{6,7}. The parallel-sequencing technique used to examine Watson's genome has also identified more than a hundred somatic rearrangements in lung cancer biopsies, and shows what is currently possible⁶.

Microarray-based techniques, like array-CGH (aCGH), use DNA hybridization to detect SNPs or CNV regions of a few thousand to more than a million base pairs⁸. These non-structural, non-sequencing methods do not supply information on genome location, so some variations, such as rearrangements, are not detected. Still, a technique like aCGH can survey the genome

more broadly than can traditional, non-parallel sequencing-based techniques. A recent aCGH study from 8 investigators led by my laboratory compared two commonly used human ES cell lines and identified nine regions with confirmed CNV differences plus 70 likely differences⁹. Though the effects of these differences have not been characterized, they could reasonably influence how cells differentiate or function.

What variations should we look for?

Of course, too-detailed an analysis of stem cell genomes could be viewed as merely a technical exercise because we do not yet know how to pluck useful information from the enormous amounts of data that would be generated. So what is the best way to analyse genomes of pluripotent stem cell lines?

Sequencing whole genomes over multiple passages, although most accurate, is impracticable. Still, to answer basic questions about genome stability in culture, as well as after genome modification, reprogramming and differentiation, high-resolution sequencing could be required. Perhaps high-density array platforms for SNP and CNV analysis, coupled with traditional karyotyping, would identify genomes best suited for clinical applications.

So far, this is easier said than done. After all, findings from genome-wide association studies in whole organisms may not be useful for stem cells. Most SNPs and CNVs associated with diabetes have yet to be linked to clinical traits such as insulin production even in the intact organism. It is too early to say that genetic variations protecting individuals from diabetes would yield clinically successful cell products. For developing treatments, currently unmeasured traits may come into play, such as how stably or efficiently transplanted cells respond to cues in their new environment. Indeed, genetic variation within transplant recipients beyond potential immune incompatibilities, rather than within transplanted cells, could become an important consideration.

Long lists of important but difficult questions should not stop scientists from considering what is possible now. A near-term consideration could be targeted array platforms that screen for those genome alterations that are strongly linked to disease susceptibilities. Another near-term consideration is planning for the future.

The scientific community has the tools for more rigorous assessments than are currently made. How to apply those tools is undoubtedly a complex problem, and at this point a solution is far from clear. Certainly, though, a first step is to get discussion of this issue onto the agenda.

[Top of page](#)

References

1. Osafune, K. *et al.* Marked differences in differentiation propensity among human embryonic stem cell lines. *Nature Biotechnol.* **26**, 313–315 (2008). | [Article](#) |
2. Wu, H. *et al.* Integrative genomic and functional analyses reveal neuronal subtype differentiation bias in human embryonic stem cell lines. *Proc. Natl Acad. Sci. USA* **104**, 13821–13826 (2007). | [Article](#) | [PubMed](#) | [ChemPort](#) |
3. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008). | [Article](#) | [PubMed](#) | [ChemPort](#) |
4. Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005). | [Article](#) | [PubMed](#) | [ISI](#) | [ChemPort](#) |
5. Draper, J. S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nature Biotechnol.* **22**, 53–54 (2004). | [Article](#) |
6. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008). | [Article](#) |
7. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005). | [Article](#) | [PubMed](#) | [ISI](#) |
8. Pinkel, D. *et al.* High-resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet.* **20**, 207–211 (1998).
9. Wu, H. *et al.* Copy number variant analysis of human embryonic stem cells. *Stem Cells* **26**, 1484–1489 (2008).

Michael Teitell is in the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at the University of California Los Angeles School of Medicine, Los Angeles, California.

e-mail: mteitell@ucla.edu www.teitell-lab.com

Featured expert



About Michael Teitell

Michael Teitell studies the mitochondrial function and genetic variation of pluripotent stem cells at the Broad Stem Cell Research Center of the University of California, Los Angeles. His lab works to understand stem cell processes, such as metabolism control and self-renewal, to determine whether and how these processes are usurped during malignant transformation.

Read the featured expert's commentary

- [Michael Teitell: How to assess a stem cell genome](#)

Featured expert's recommendations

- [Characterization of human embryonic stem cell lines by the International Stem Cell Initiative](#)

This paper describes an international consortium's initial characterization of 75 human embryonic stem cell (hES cell) lines and establishes each line's unique identity.

Adewumi, O., *et al.* *Nature Biotechnol.* **25**, 803-816 (2007).

- [Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*](#)

Andrews and colleagues describe hES cell line genome instability caused by culture adaptation over time, with many changes similar to those seen in human tumours - especially testicular germ cell cancers.

Baker, D.E., *et al.* *Nature Biotechnol.* **25**, 207-215 (2007).

- [Copy number variant analysis of human embryonic stem cells](#)

This paper examines copy number differences between two hES cell lines that differ in their neurogenic potential as one putative source for varying functional and differentiation outcomes.

Wu, H. *et al.* *Stem Cells* **26**, 1484-1489 (2008).

- [Marked differences in differentiation propensity among human embryonic stem cell lines](#)

Melton and colleagues describe variations in the differentiation potential of 17 hES cell lines that could arise from genome differences in outbred human populations as well as from varying epigenetic states.

Osafune, K. *et al.* *Nature Biotechnol.* **26**, 313-315 (2008).

- [Global variation in copy number in the human genome](#)

One of several articles exploring the extent and variability of copy number variation in somatic

cells from human subpopulations, which by extension suggests the amount of genome complexity between hES cell lines from unrelated donors.

Redon, R. *et al. Nature* **444**, 444-454 (2006).

- [Genomics: Massively parallel sequencing](#)

This brief commentary by Rogers and Venter introduced massively parallel sequencing as a viable approach for determining genome-wide genetic heterogeneity.

Rogers, Y. H. & Venter, J. C. *Nature* **437**, 326-327 (2005).

- [TORC2 regulates germinal center repression of the TCL1 oncoprotein to promote B cell development and inhibit transformation](#)

This paper describes a previously unknown mechanism for regulating TCL1, a lymphocyte oncogene that regulates mouse ES cell self-renewal, and provides a potential connection between malignant transformation and mechanisms of pluripotent stem cell maintenance.

Kuraishy, A. I. *et al. Proc. Natl. Acad. Sci. USA* **104**, 10175-10180 (2007).