

Supplementary Table 1 - Examples of different terms mapping to tissue diagnosis describing Adenocarcinoma: Cell line diagnosis terms from public databases are shown. Original cell line name and diagnosis are listed along with the curated cName, Primary Tissue and Tissue Diagnosis Terms.

Supplementary Table 2 - Annotations for 3,587 cell lines: Manually curated list of cell line names and associated descriptive data using controlled vocabulary. “Unresolved” indicates fields where insufficient or conflicting information could not be easily resolved. “Unknown” indicates information was not found/not available. A single source was listed for each cell line, however other sources may be available. Site of extraction is only listed when it is a site other than the primary tissue. No longer available indicates the cell line was provided by a cell repository, but has since been removed from distribution.

Supplementary Table 3 - STR reference file: A curated list of SNPs linked to cName and Synonymous cell lines. This table contains a list of unique STRs. Profiles contain at least 8 loci. For Synonymous, Misidentified and derivative cell lines with identical STR profiles, a single profile is listed with the names of all cell line mapped to the “Synonymous, Misidentified and derivative cell lines” column. In cases where profiles of the same cell line differ slightly, these profiles are retained (e.g. 293 from ATCC and DSMZ).

Supplementary Table 4 - STR analysis of cell line SK-N-BE(2): The STR profile from SK-N-BE(2) (ATCC CRL-2271, “Query”) was compared to our (“GNE matches”), ATCC (“ATCC Matches”) and DSMZ (“DSMZ matches”) databases. Hits greater than 0.8 EV (evaluation value) for DSMZ and 80% match for ATCC are shown in comparison to our percent match analysis (see Methods). Differences between the query sample alleles and matched STR profiles are highlighted in grey.

Supplementary Table 5a - SNP base calls reference file: A curated list of 1,021 SNP base calls linked to cName. Cell line identity were validated by STR profiling.

Supplementary Table 5b - SNP XY calls reference file: A curated list of 1,021 SNP XY calls linked to cName. Cell line identity were validated by STR profiling.

Supplementary Table 6 (Synonym Table) - A list of synonymous cell lines: These are cell lines proven to be of common origin (i.e. matching DNA profiles). This table includes legitimate synonymous lines (derived from the same patient or from a common parental cell line), misidentified cell lines and lines where the relationship is unknown. Associations were primarily determined by comparison of STR and SNP profiles and from published sources (see Methods). A comment is included regarding the status of each cell line e.g. where the cell line is believed to be parental, derivative or contaminated/misidentified. Comments are based on available information and are subject to revision. It is strongly advised to always double-check the provenance of any cell line prior to use.

Supplementary Table 7 - List of misreported synonymous cell lines: A list of cell lines which have been reported as having the same STR profile, for which we found evidence of unique profiles. Cell lines may be listed as misreported for several reasons. Authentic stocks may exist elsewhere, or more than one cell line may exist with the same name.

Supplementary Table 8 - Cancer Cell Line Encyclopedia (CCLE) cell lines and annotation: CCLE cell lines annotated with synonymous, misidentified and derivative cell lines, overlap with Sanger cell lines, comparison of cell line naming and curated cName, primary tissue and tissue diagnosis terms.

Supplementary Table 9 - Sanger cell lines and annotations: Sanger cell lines annotated with synonymous, misidentified and derivative cell lines, overlap with CCLE cell lines, comparison of cell line naming and curated cName, primary tissue and tissue diagnosis terms.

Supplementary Table 10 - Pairwise analysis of STR profiles in Table S3.

Supplementary Table 11 - Controlled vocabularies for Tissue

Supplementary Table 12 - Controlled vocabularies for Tissue Diagnosis

Supplementary Table 13 - Controlled vocabularies for Tissue Ontologies

Supplementary Table 14 - Controlled vocabularies for Tissue Diagnosis Ontologies