GXB_Clinical_Variables

15 March 2016

Find all unique variable names in all datasets

```
ALL_VARIABLES <- c(names(BOR), names(EMC1), names(EMC2), names(EMCT), names(GUYT2), nam
es(MAINZ), names(STO), names(MSKCC), names(YAU1), names(PMC), names(PMC3), names(TBIG),
names(YAU2), names(UPSI))
UNIQUE_VARS <- unique(ALL_VARIABLES)
```

Change all dubble dots to single dots and change every letter to upper case, find unique variables

```
UNIQUE_VARS <- gsub("\\.\\.", ".", UNIQUE_VARS)
UNIQUE_VARS <- toupper(UNIQUE_VARS)
unique(UNIQUE_VARS)
```

```
##   [1] "SAMPLE.ID"
##   [2] "ARRAY.SAMPLE.ID"
##   [3] "BARCODE"
##   [4] "X.SAMPLE.TITLE"
##   [5] "DSS.TIME"
##   [6] "ADJUVANT.TREATED..0.NO.1.YES."
##   [7] "CHARACTERISTICS"
##   [8] "CHEMO..0.NO.1.YES."
##   [9] "DFS.EVENT.DEFINED.AS.ANY.TYPE.OF.RECURRENCE.OR.DEATH.FROM.BREAST.CANCER."
##  [10] "DFS.TIME"
##  [11] "DMFS.EVENT.DEFINED.AS.DISTANT.METASTASIS.OR.DEATH.FROM.BREAST.CANCER."
##  [12] "DMFS.TIME"
##  [13] "DSS.EVENT.DEFINED.AS.DEATH.FROM.BREAST.CANCER."
##  [14] "ER.STATUS"
##  [15] "HER2.STATUS"
##  [16] "HISTOLOGICAL.GRADE"
##  [17] "HISTOLOGY"
##  [18] "LN.STATUS"
##  [19] "PATIENT.AGE"
##  [20] "PGR.STATUS"
##  [21] "SIZE.MM."
##  [22] "TREATMENT.TYPE"
##  [23] "IMS"
##  [24] "DMFS.10Y.TIME"
##  [25] "DMFS.10Y.EVENT"
##  [26] "IBS"
##  [27] "IDS"
##  [28] "ICR"
##  [29] "BONE.RELAPSES.1.YES.0.NO."
##  [30] "ID"
##  [31] "COHORT"
##  [32] "LMS.STATUS"
##  [33] "LUNG.MET.ALL"
##  [34] "LUNG.MET.FIRST.EVENT"
##  [35] "LUNG.MET.FREE.SURVIVAL.YR."
##  [36] "METASTASIS"
##  [37] "METASTASIS.FREE.SURVIVAL.YR."
##  [38] "DFS"
##  [39] "BREAST.CANCER"
##  [40] "AGE"
##  [41] "E"
##  [42] "ER"
##  [43] "GRADE"
##  [44] "NODE"
##  [45] "PGR"
##  [46] "SAMPLENAME"
##  [47] "SERIES"
##  [48] "SIZE"
##  [49] "T"
##  [50] "TREATMENT"
##  [51] "CLINIC"
##  [52] "SIZE.IN.CM"
##  [53] "STORAGE"
```

```
##  [54] "BREAST.TUMOR.TISSUE"
##  [55] "DEATH"
##  [56] "DEATH.BC"
##  [57] "ELSTON"
##  [58] "RELAPSE"
##  [59] "SUBTYPE"
##  [60] "SURV.DEATH"
##  [61] "SURV.RELAPSE"
##  [62] "SEE.SERIES.GSE1456.RECORD.FOR.ADDITIONAL.INFORMATION"
##  [63] "AGE.AT.DX"
##  [64] "ALL.LYMPH.NODES"
##  [65] "BINARY.5Y.MET.EVENT"
##  [66] "BM.EVENT"
##  [67] "BMFS.YR."
##  [68] "LM.EVENT"
##  [69] "LMFS.YR."
##  [70] "MET.EVENT"
##  [71] "MFS.YR."
##  [72] "NAME"
##  [73] "PATH.ER.STATUS"
##  [74] "PATH.PR.STATUS"
##  [75] "POS.LYMPH.NODES"
##  [76] "TISSUE.TYPE"
##  [77] "TUMOR.SIZE.CM."
##  [78] "VAN.T.VEER.SIGNATURE"
##  [79] "ESTROGEN.RECEPTOR.ER."
##  [80] "NODAL.STATUS"
##  [81] "PROGESTERONE.RECEPTOR.PR."
##  [82] "PROGESTERONE.RECEPTOR.PR..1"
##  [83] "DISTANT.RFS"
##  [84] "GGI"
##  [85] "TIME.RFS"
##  [86] "AOL.OS.10Y"
##  [87] "ANGIOINV"
##  [88] "HISTTYPE"
##  [89] "LYMP.INFIL"
##  [90] "NPI"
##  [91] "SURGERY.TYPE"
##  [92] "FILENAME"
##  [93] "HOSPITAL"
##  [94] "RISK.AOL"
##  [95] "RISKNPI"
##  [96] "RISKSG"
##  [97] "VERIDEX.RISK"
##  [98] "DFS.YEARS"
##  [99] "ER."
## [100] "RECURRENCE.1.EVENT.0.CENSORED."
## [101] "DMFS10YTIME"
## [102] "DMFS10YEVENT"
## [103] "ALL.PATIENTS.1.INCLUDED.IN.SURVIVAL.ANALYSIS."
## [104] "DFS.EVENT.0.CENSORED.1.EVENT.DEFINED.AS.ANY.TYPE.OF.RECURRENCE.LOCAL.REGIONA
L.OR.DISTANT.OR.DEATH.FROM.BREAST.CANCER"
## [105] "DFS.TIME.YRS."
## [106] "ER..ENDOCRINE.THERAPY.ONLY.1.INCLUDED.IN.SURVIVAL.ANALYSIS."
```

```
## [107] "ELSTON.NGS.HISTOLOGIC.GRADE"
## [108] "GENETIC.GRADE.SIGNATURE.STATUS.PREDICTION.BY.SWS.CLASSIFIER."
## [109] "INDEX.ID."
## [110] "LYMPH.NODE.STATUS"
## [111] "NUH.SAMPLE.ID"
## [112] "NO.SYSTEMIC.THERAPY.1.INCLUDED.IN.SURVIVAL.ANALYSIS."
## [113] "PROBABILITY.1.LIKE.BY.SWS.CLASSIFIER."
## [114] "PROBABILITY.3.LIKE.BY.SWS.CLASSIFIER."
## [115] "AGE.AT.DIAGNOSIS"
## [116] "P53.SEQ.MUT.STATUS.P53.MUTANT.P53.WILDTYPE."
## [117] "TUMOR.SIZE.MM."
```