gbg.HTG.normalize.v1

This function calculates normalized gene expression values according to “Normalizing\_HTG\_data\_V2.0.pdf” (equation 3 on page 10) from data from the HTG EdgeSeq Oncology Biomarker Panel.

The HTG data must be count values. Positive controls, negative controls and PCLs (alien probe sets with prefix “ER-“) must be omitted; housekeeping genes must be included; the total number of genes (probe sets) is 2549.

R versions: V3.3.2

Libraries: none

Dependencies: none

Owner: Karsten Weber

Parameters:

| **Name** | **Values** | **Default** | **Description** |
| --- | --- | --- | --- |
| counts.mat | integer matrix, ncol=2549 | <must be specified> | Count values (gene expressions) from the HTG EdgeSeq Oncology Biomarker Panel.  Each matrix column corresponds to a gene (probe set); column names should equal the probe set names as defined by the HTG panel (names are not checked, because there may be some variations, e.g. "HLA-A" or "HLA\_A"). Positive controls, negative controls and PCLs (alien probe sets with prefix “ER-“) must be omitted; housekeeping genes must be included.  Each matrix row corresponds to a sample. Row names are used to identify samples: They are contained in the returned matrix (see below).  The gene expression data should have passed quality control (QC) checks; one of them ensures a required minimum of total counts for a sample (this may include controls and PCLs, so cannot be calculated from counts.mat). |

Returns:

The function returns a double matrix:

* The number, order, and names of rows (samples) are the same as in counts.mat.
* The number, order, and names of columns (genes) are the same as in counts.mat.
* Each matrix element is a normalized gene expression value: a real number >= 3.

Example:

d.raw <- read.csv2("G:/Statistik/Bioinformatik/PenelopeB/HTG/HTG Daten aus idgard/20200413\_Penelope\_Lauf\_1\_parsed\_data\_QCed\_Raw.csv", stringsAsFactors = F, skip = 9)

d.raw <- subset(d.raw, !grepl("^Total Counts$|^NEG\_CTRL\_|^POS\_CTRL\_|^ER-|^$", Sample.Name))

d.counts <- t(as.matrix(d.raw[, -1]))

colnames(d.counts) <- d.raw$"Sample.Name"

d.norm <- gbg.HTG.normalize.v1(d.counts)