gbg.HTG.AIMS.v1

This function calculates the AIMS subtype (Paquet & Hallett, JNCI J Natl Cancer Inst 107, 2015) from data from the HTG EdgeSeq Oncology Biomarker Panel.

The HTG data must be count values (not normalized 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, see gbg.HTG.normalize.v1 or “Normalizing\_HTG\_data\_V2.0.pdf” for details.

When calculating the AIMS subtype all genes are considered that are used by AIMS and contained in the HTG panel; in total 91 genes and 42 rules are used. The calculation is performed as Pfizer did for the PALOMA-3 data (as a reference).

R versions: V3.3.2

Libraries: AIMS (V1.6.0)

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 (only the names of the 91 genes used by AIMS need to be present). Positive controls, negative controls and PCLs (alien probe sets with prefix “ER-“) should be omitted; housekeeping genes should be included.  Each matrix row corresponds to a sample. Row names are used to identify samples: They are contained in the returned data.frame (see below). |
| published.names | logical scalar | T | Determines which kind of names for the five AIMS subtypes are used in the data.frame returned. Value TRUE uses subtype names used in the publication (Paquet & Hallett, JNCI J Natl Cancer Inst 107, 2015) (table 2); value FALSE uses subtype names as returned by the R package “AIMS” (same authors). See also below. |

Returns:

The function returns a data.frame with each row corresponding to a sample and with the following columns:

| **Name** (published.names=F) | **Name** (published.names=T) | **Values** | **Description** |
| --- | --- | --- | --- |
| ID | | character | Sample identifier. These are the row names of parameter counts.mat (same order). |
| AIMS\_class | | character | AIMS subtype with the highest posterior probability.  The AIMS subtype name also depends on parameter. published.names. |
| AIMS\_p\_Basal | AIMS\_p\_BasalL | double | Posterior probabilities of each AIMS subtype. The five numbers are from the interval [0, 1] and sum up to 1. |
| AIMS\_p\_Her2 | AIMS\_p\_HER2E | double |
| AIMS\_p\_LumA | AIMS\_p\_LumA | double |
| AIMS\_p\_LumB | AIMS\_p\_LumB | double |
| AIMS\_p\_Normal | AIMS\_p\_NormL | double |

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"

print(gbg.HTG.AIMS.v1(d.counts))