# Sample ID format for NGS data (v5.33)

2022_11_17

Lei Wei

# Full sample ID

Old

BRC001-T(N)d(rmcxbavp)1a-ESc1_RS-01516230-RS-01516231_PT-00090080_WGS-ILU

- BRC001-Td1a (publication ID)

**Cohort and patient**

BRC : publication cohort name, to be determined by PI. For sequential projects of same disease, it is recommended to distinguish them by letter ("a,b,c") instead of number to avoid confusion with *patient publication number*.

001 : patient publication number, to be determined by the PI

**Sample type and source**

T : primary tissue type, T (tumor or test) or N (normal)

d : tissue type source, d(primary diagnosis), r(relapse), m(metastasis), c(cell line), x(xenograft), b(blood), a(adjacent normal tissue) , n(non-adjacent normal tissue) , ~~s(stem cell),~~ v(saliva), p(plasma), f(fluid), t(treatment). Can have multiple, e.x. cx (cell line derived xenograft), s(simulated), o(organoid)

1 : if there are multiple tumors/normals, 1, 2, 3…

a : if a tumor contains multiple sites, a, b, c… Extra digits are allowed (e.x. a1, a2..)

Section d1a is optional depending on design of publication.

**Additional procedure (optional)**

An extra tag ("-ESc1" in the example) can be added to label additional procedure(s) performed to the sample. The tag uses capital letter(s) to indicate the type of procedure, and lower letter(s) and/or number(s) as flexible identifier.

ES : additional procedures. Currently allow: C(culture, usually include cutting up the tissue into tiny pieces), D(dissociated spheres), E(stem cells), S(single cells, e.x. by C1 system), T(treatment), F(fluorescence), K(spike in other sample), U(UMI), R2(rerun #2). When there are multiple procedures, the letters are ordered by alphabetical order to ensure uniqueness.

c1 : flexible identifier which is optional. The PI can use any combination of lower letter(s) and/or number(s).

- RS-01516230
  - 2014_01_03: if there are multiple libraries, then all RS ids will be shown in this field, separated by dashes (e.x. RS-01516230-RS-01516231)
- PT-00090080 (patient ID or cell line name for tracking across multiple diseases)
- WGS-ILU (sequencing platform)
  - WGS, WES, RNA, CUS (customized capture sequencing)
  - ILU (Illumina), ION (Ion Torrent), SOD (Solid), 454, PAC (Pac bio) …
  - In publications to compare multiple platform, it is legal to use sequencing platform as a 'Additional procedure' (see above) to differentiate samples. E.x. "EVNC001-Td1-WES" vs "EVNC001-Td1-CUS"

# Overview of project

| | Purpose | Samples | Conditions | Sample |
|---|---|---|---|---|
| aim1 (human wide) | establish narrow panel | 375 samples from 25 donors (discover) + 225 sample from 15 donors(validation) | - | normal epidermis SE/NE, dermis |
| mouse tumor | establish wide panel | SKH1 CSCC tumors | - | CSCC T/N |
| aim1(mouse wide) | establish narrow panel | 480 samples from 8 mice | UVB Doses(4) | normal epidermis SE/NE, dermis |
| aim1K&2(human narrow) | kinetics by age | 1160 samples from 40 donors(discover)+580 samples from 20 donors(valid) | Age groups (4) | CSCC T/N, epidermis SE, dermis |
| aim1K(mouse narrow) | kinetics by dose | 208 mice x 8 samples/mice | UVB Doses(3)/Weeks(5), before/after FT | CSCC T/N, epidermis SE, dermis |
| aim 2(human narrow) | CM vs cancer risk | 408 samples from 17 pts x 3 + validation: 80 samples | high/low risks/BCC not CSCC | CSCC T/N, epidermis SE, dermis |
| aim2(mouse narrow) | CM vs cancer risk | 70 mice x 8samples/mice | UVB Doses(11)/Weeks(1) | CSCC T/N, epidermis SE, dermis |
| aim 3 (mouse narrow) | field treatment efficacy | 192 mice x 8 samples/mice | ALA(2), 5FU(2), control | CSCC T/N, epidermis SE, dermis |
| aim 3 (human narrow) | field treatment efficacy | 300 samples | before/after FT | CSCC T/N, epidermis SE, dermis |
| aim 4 (mouse narrow) | photoprotection | 160 mice/880 samples | UVB Dose(3)/Weaks(3)/SPF(3)weeks(3) | CSCC T/N, epidermis SE, dermis |

# Potential types of information to include

| **Potential conditions** | **add to ID** |
|---|---|
| human/mouse | cohort ID (h/m) |
| Primary tissue type | S (skin), B (Bladder), |
| Secondary tissue type | p (epidermis) vs d (dermis) |
| Third tissue type | e (sun-exposed) vs n (non-exposed) |
| Biopsy # | 1-? |
| UVB doses | uv,1-11 ("uv1") |
| weeks | w, ("w1") |
| field treatment (ALA) | fta, ("fta11"-level1 before, "fta22"-level 2 after) |
| field treatment (5FU) | ftu, ("ftu11"-level1 before, "ftu22"-level 2 after) |
| sun screen | ss, ("ss11"-level1 before, "ss22"-level 2 after) |
| risks | - |
| wide/narrow sequencing | sequencing platfrom(UTSw/UTSn) |
| discovery/validation | - |

Pub ID NGS ID

**SCMm0001-Spe008-Us125w2Fu125w06Sm05w06Mc15w10_UTSn_PT#_RS#(s).bam**

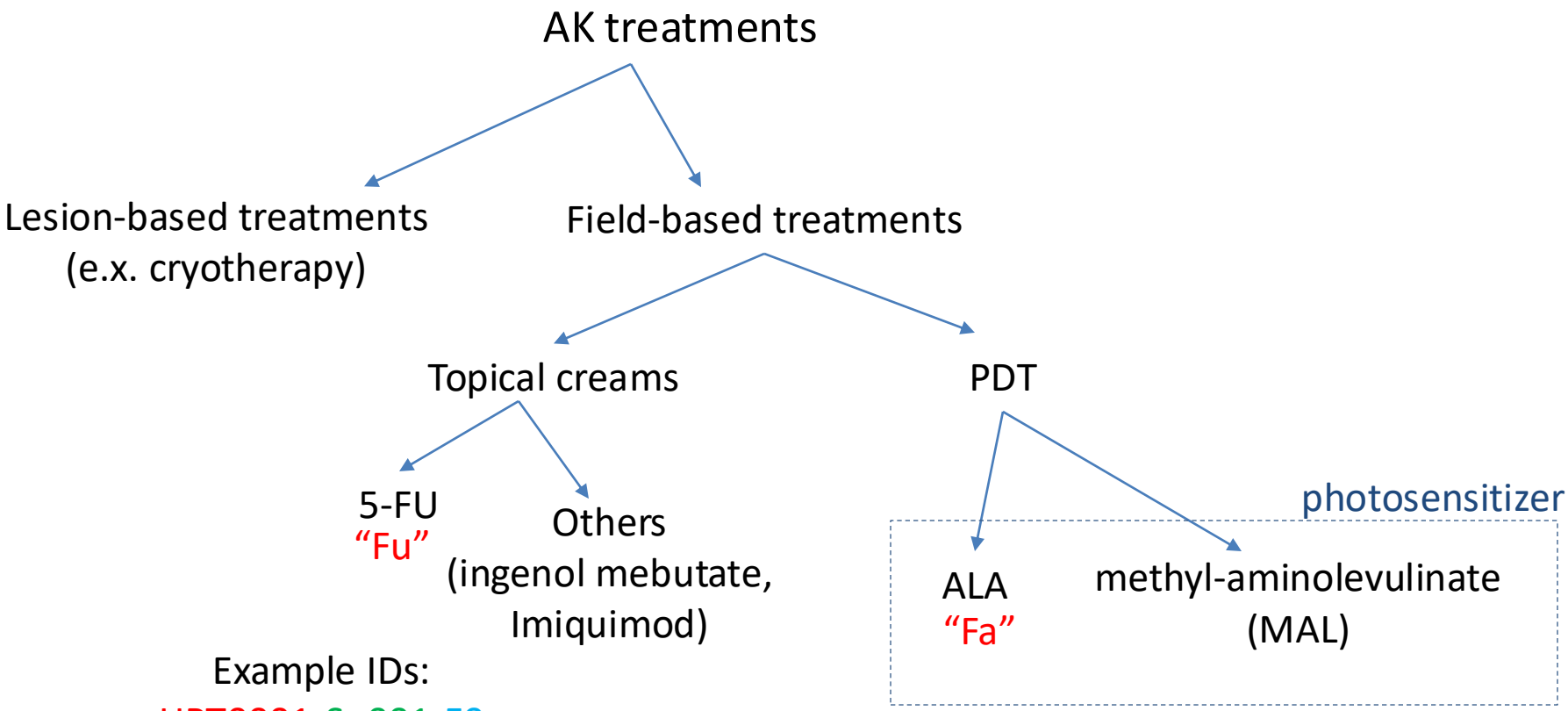
Capitalize – primary variable

fixed

flexible by project

| | Information | Examples | Letter |
|---|---|---|---|
| **Patient** | Project | SCM (skin clonal mutation) | CAPITAL |
| | Species | h-homo sapcience; m-mous | lower |
| | Donor | 0001 (4-digits) | numeric |
| **Sample** | Primary tissue site | S (skin), P (prostate) | CAPITAL |
| | Secondary tissue type | p (epidermis, normal) vs d (dermis, normal) vs t (tumor, clinically), vs pz for peripheral zone and tz for transitional zone (prostate) | lower |
| | Third tissue type (area) | e (sun-exposed) vs n (non-exposed) | lower |
| | Biopsy # | 015 (3-digits, multiple biopsies of the same area) | numeric |
| **Treatment** | format: type (**capital**/lower letters) followed by value (numeric, as deliminator) | | |
| | **U**V irritiation type with dose (uvb or uva or ss; 3digits for dose) | Ub (Uvb); Ua(Uva); Us (Us solar simulator),e.x. Us125 | Capital first letter + lower+numeric |
| | Treatment weeks (2 digits) | w, ("w06") | lower+numeric |
| | **F**ield treatment (needs to be followed by type and length except for F0) | F0-no field treatment; Fa (ALA); Fu (5FU); Fc (carevdilol), F (combination?) | Capital first letter + lower+numeric |
| | field treatment length (weeks) | w, ("w06") | lower+numeric |
| | **S**pf: sun screen, types? | "m" for mock sun-screen, or "s" for sun-screen, followed by a 2-digit numeric (e.x. 05), "Ss05" | lower+numeric |
| | sun screen length (weeks) | w, ("w06") | lower+numeric |
| | I**m**munosuppression | "c" for cyclosporine, "a" for azathioprine, "m"-mycophenolate mofetil, "s"-sirolimus, dose-numeric, e.x. "Mc15" | lower+numeric |
| | length (weeks) | w, ("w10") | lower+numeric |
| **NGS** | wide/narrow sequencing | sequencing platform(UTSw/UTSn/WES) | CAPITAL+lower |

# Question for field treatments




Example IDs:

HPT0001-Sp001-F0

HPT0001-Sp005-Fa

Future: Fc (carevdilol), F (combination?)