

```
function [Data_Filtered] = GA_OccurrenceFilter(Data_Matrix,Groups,A_thres,G_thres)
```

```
% Data_Matrix is a proteomics dataset with orientation, rows = samples and columns equal peptides

% Groups is a vector of class indices associated with each observation in Peptide_Vector. This vector
% is numeric and sequential, which means that if there are 4 classes, then each position in
% Peptide_Vector will be represented in "Groups" as a 1,2,3 or a 4. If the data are in another
% format, such as a Cell structure it would need to be converted.

% A_thres is the number of samples that must be observed in at least 2 groups to pass the ANOVA-
% filter (2 or 3 typically).

% G_thres is the significance value required to identify the minimum number of observations required
% in a single group (G-filter): (0.05 in paper)

% 1) Determine how many samples are in each group (nk in paper)
N = length(Groups);
num_g = max(Groups);
nk = zeros(num_g,1);
for i = 1:num_g
    nk(i) = length(find(Groups == i));
end

% 2) Determine how many samples must have observations to pass the G-filter
G_FilterNumber = IP_GFilter(Groups,num_g,nk,G_thres);

% 3) Identify which peptides pass either a G-filter or an ANOVA-filter
[~,P] = size(Data_Matrix); % N = number of samples, P = number of peptides
IDX = zeros(P,1);
A_FilterNumber = repmat(A_thres,1,num_g);
for i = 1:P
    np = NaN(1,num_g);
    for j = 1:num_g
        np(j) = nk(j) - sum(isnan(Data_Matrix(Groups == j,i)));
        % the number present in each group
    end
    tmp = np - G_FilterNumber;
    if max(tmp) >= 0
        IDX(i) = 1; % at least 1 peptide had enough data for the G-filter
    else
        tmp = np - A_FilterNumber;
        if length(find(tmp >=0)) > 1
            IDX(i) = 1; % e.g., at least 2 peptides have enough data for the ANOVA-filter
        end
    end
end
end

% 4) Filter the peptide to only those that pass at least one of the two filters
Data_Filtered = Data_Matrix(:,IDX == 1);
end
```

**Figure S6.** Generic MatLab® code for performing the GA-Filter based on an ANVOA number (A\_thres) and the number of identifications associated with the G-Filter. This code depends upon the code described in Figure S5.