## Supplementary Information:

**Supplementary Information 1:**

The statistical modeling in this study is based on Bayesian networks. Bayesian networks are directed acyclic graph structures that extend Bayesian analysis11, and are a set of multivariate probabilistic models that have increased power in learning and classification due to their compact factorization of data2,3. 2Bayesian networks are powerful in their ability to learn conditional relationships from large datasets and to use this probability distribution to classify other instances based on their feature values. When they are used to represent biological systems (Fig. S-1),Bayesian networks create models of simultaneous genetic associations and dependencies, as well as genetic interplay with clinical and environmental variables3.

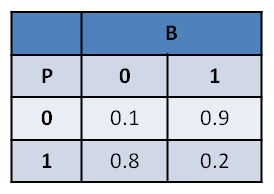
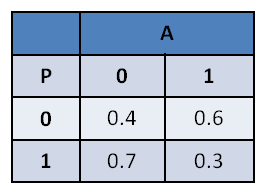
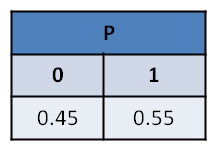


Figure S - 1: *Bayesian Network Applied to Biological Systems*. *P* is the class variable, *i.e.*, phenotype, while *A*, *B*, and *C* are features, *i.e.* genes. Each edge represents either a gene-phenotype or gene-gene association.

**Supplementary Information 2:**

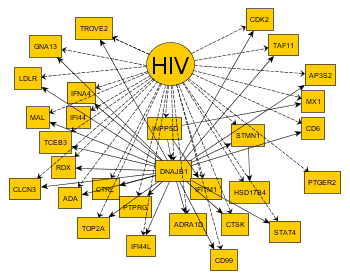


Figure S-2: *Singly-Structured Bayesian Classifier for HIV.* This single TAN is only able to classify HIV-infected patients with an AUROC of 0.851.

**Supplementary Information 3:**

Table S - 1: *GEO Datasets for HIV, Kaposi’s Sarcoma, and Cancers of the Blood*. The following GDS were taken from the Gene Expression Omnibus to analyze the relationship between disease and gene expression.

|  |  |  |
| --- | --- | --- |
| **HIV Infection** | **Kaposi’s Sarcoma** | **Blood Cancers** |
| GDS1449 | GDS940 | GDS2908 |
| GDS171 | GDS988 | GDS3516 |
| GDS1726 | GDS1063 | GDS3057 |

Table S-2: *Results of* *prediction-based* *Enrichment Analysis for HIV Infection.* Enrichment analysis of the shared-feature set in reveals GO and KEGG biological concepts related to HIV infection.

|  |  |
| --- | --- |
| **Biological Concept** | ***p-*value** |
| Mitotic cell cycle | 0.009 |
| Cell cycle process | 0.009 |
| Organelle organization and biogenesis | 0.013 |
| Cellular macromolecule metabolic process | 0.019 |
| Intracellular signaling cascade | 0.027 |
| Phosphoric monoester hydrolase activity | 0.042 |
| Signal transduction | 0.057 |
| B cell receptor signaling pathway | 0.068 |

Table S - 3: *Significant Gene List for Kaposi’s Sarcoma and Blood Cancers*. The shared features for the top 10% of genes in Kaposi’s sarcoma experiments and for the top 5% of blood cancer-related genes are shown below.

|  |  |
| --- | --- |
| **Kaposi’s Sarcoma** | **Blood Cancers** |
| *SEPP1*  *BCL2A1*  *CXCR7*  *RGS1*  *WSB1*  *CXCR3*  *CEACAM1*  *APBB2*  *DICER1*  *KYNU*  *VIP*  *CDC42EP3*  *SLC1A1*  *KLF5*  *PYROXD1*  *SPTBN1*  *PIK3CD*  *LPIN2*  *ADD3*  *IDS*  *C2orf3*  *CHRM3*  *DTX4*  *CEBPD*  *PDLIM5*  *DHX57*  *FUT4* | *CD59*  *MICALL2*  *RARRES3*  *GGT1*  *SEC62*  *ZBTB43*  *XKR8*  *COL6A1*  *TPD52*  *ANK1*  *CALR*  *LSM14A*  *CORO1B*  *DNPEP*  *DYNLT1*  *PDE6G*  *DAG1*  *RASSF7* |

Table S - 4: *Cross-Validation Results for 3-Fold Cross-Validation, Kaposi’s Sarcoma and Blood Cancer Multi-net Classifiers.* The AUROC for the Kaposi’s sarcoma and blood cancer integrative and singly-structured classifiers are shown below.

|  |  |  |
| --- | --- | --- |
| **Disease** | **Multi-net** | **Singly-Structured** |
| Kaposi’s Sarcoma | 0.893 | 0.658 |
| Blood Cancers | 0.970 | 0.943 |

Table S - 5: *Genes in Significant Chromosomal Regions*. Many genes related to HIV infection, Kaposi’s sarcoma, and cancers of the blood are in the chromosomal regions listed below.

|  |  |
| --- | --- |
| **Chromosomal Region** | **Genes** |
| 1p31–36 | *IFI44*  *IFI44L*  *TCEB3*  *XKR8*  *PIK3CD* |
| 2p21–22 | *CDC42EP3*  *SPTBN1*  *DHX57* |
| 2q32–37 | *STAT4*  *PNPEP*  *KYNV*  *INPP5D*  *HSPD1* |
| 11q21–23 | *RDX*  *RARRES3*  *FVT4* |

**Supplementary Information 4:**

Because HIV is a retrovirus, the virus spreads throughout the body by integrating its own genetic information into the human genome4. Therefore, transcriptional regulation, which involves the genes *STAT4*, *TOP2A*, *TAF11*, and *TCEB3*5, is crucial to the progression of HIV infection. The Bayesian multi-net seems to reflect the important epistatic role of *STAT4*, a gene previously found to play an important immunoregulatory role in HIV patients6. The *STAT4­*–*TOP2A* link is present in the ‘PBMC, In Vivo’ sub-network of the Bayesian multi-net. The ‘HIV Encephalitis’ network reflects the epistatic interaction between *STAT4* and another transcription-regulating gene *TAF11*, which up-regulates viral transcription in the host5, making this link biologically plausible.

**Supplementary Information 5:**

Several genes in Table S-3 related to Kaposi’s sarcoma and cancers of the blood, *CEACAM1*, *DICER1*, and *KLF5*5, relate to angiogenesis, which has been associated with tumor growth7. Furthermore, *VIP*, *RARRES3*, and *CHRM3* are related to cell proliferation5, related to the progression of HIV infection8,9, while *BLC2A1*, *CXCR7*, and *CXCR3* allow tumor growth through anti-apoptosis activity5. Similarly, the down-regulation of *APBB2* in cancer patients prevents normal cell cycle arrest, allowing abnormal growth, and *DYNLT1* affects the cycle through its functionality in forming the mitotic spindle5. Over-expression of the gene *TPD52*, coding for tumor protein D52, has also been implicated in many cancers in the past10,11. Two blood cancer-related genes also reflect blood function, making these significant genes biologically plausible: *CD59* is involved in blood coagulation while *PDE6G* is involved in platelet activity5. For more discussion on the network constructed for Kaposi’s sarcoma and blood cancers see Supplementary Information 4.

The multi-nets constructed for Kaposi’s sarcoma (Figure 2) and cancers of the blood (Figure 3) outperform the singly-structured models, as shown in Table S-4 (Supplementary Information). While the Kaposi’s sarcoma multi-net outperforms the singly-structured model (AUROC = 0.893 vs. 0.658), the difference between the performance of the multi-net and single TAN is only 0.027 (AUROC = 0.970 vs. 0.943). The presented multi-net classifier for blood cancers is also the first of its kind to integrate both lymphoma and leukemia patients into one model, showing that patients with different cancer types can be classified reliably using a common set of genes. Furthermore, these supplemental analyses show that the integrative approach presented in this study is generalizable to other diseases besides HIV/AIDS.

**Supplementary Information 6:**

## chromosomes.jpg

Figure S-3: *Significant Chromosomal Regions.* Gene expression in the four bracketed regions on Chromosomes 1, 2, and 11 is significant to HIV and HIV-related cancers. These regions each encompass at least three genes from Table 1 and Table S - 3. The sizes of respective chromosomes are not to scale.

References:

1. Pearl,J. (1988) *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, New York.
2. Alterovitz,G., Liu,J., Afkhami,E., and Ramoni,M.F. (2007) Bayesian methods for proteomics. *Proteomics*, **7**, 2843–2855.
3. Sebastiani,P., Ramoni,M.F., Nolan,V., Baldwin,C.T., and Steinberg,M.H. (2005) Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat. Genet.*, **37**, 435–440.
4. Schroder,A.R.W., Shinn,P., Chen,H., Berry,C., Ecker,J.R., and Bushman,F. (2002) HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell*, **110**, 521–529.
5. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C., *et al*. (2004) The Gene Ontology (GO) database and informatics re­source. *Nucleic Acids Res*., **32**, D258–261.
6. Watford,W.T., Hissong,B.D., Bream,J.H., Kanno,Y., Muul,L., and O’Shea,J.J. (2004) Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. *Immunol. Rev.*, **202**, 139–156.
7. Carmeliet,P. and Jain,R.K. (2000) Angiogenesis in cancer and other diseases. *Nature*, **407**, 249–257.
8. Migueles,S.A., Laborico,A.C., Shupert,W.L., Sabbaghian,M.S., Rabin,R., Hallahan,C.W., Van Baarle,D., Kostense,S., Miedema,F., McLaughlin, *et al*.(2002) HIV-specific CD8+ T cell proliferation is coupled to perform expression and is maintained in nonprogressors. *Nat. Immunol.*, **3**, 1061–1068.
9. McNeil,A.C., Shupert,W.L., Iyasere,C.A., Hallahan,C.W., Mican,J.A., Davey,R.T. Jr., and Connors,M. (2001) High-level HIV-1 viremia suppresses viral antigen-specific CD4+ T cell proliferation. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 13878–13883.
10. Rubin, M.A., Varambally,S., Beroukhim,R., Tomlins,S.A., Rhodes,D.R., Paris,P.L., Hofer,M.D., Storz-Schweizer,M., Kuefer,R., Fletcher,J.A., *et al*. (2004) Overexpression, Amplification, and Androgen Regulation of TPD52 in Prostate Cancer. *Cancer Res.*, **64**, 3814–3822.
11. Byrne,J.A., Balleine,R.L., Fejzo,M.S., Mercieca,J., Chiew,Y.-E., Livnat,Y., St. Heaps,L., Peters,G.B., Byth,K., Karlan, B.Y., *et al*. (2005) Tumor protein D52 (TPD52) is overexpressed and a gene amplification target in ovarian cancer. *Int. J. Cancer*, **117**, 1049–1054.